EDA - Exploratory Data Analysis

importing data

```
Loaded df_appearances from ../pickles/df_appearances.pkl
Skipping player_performance...
Skipping player_game_team_mapping...
Skipping df_games_odds...
Loaded df_teamstats from ../pickles/df_teamstats.pkl
Loaded df_shots from ../pickles/df_shots.pkl
Loaded gameresult from ../pickles/gameresult.pkl
Loaded df_after_outliers_missing from ../pickles/df_after_outliers_missing.pkl
Skipping team_performance...
Loaded df_with_categories from ../pickles/df_with_categories.pkl
Loaded df_num_after_EDA from ../pickles/df_num_after_EDA.pkl
Skipping df_games...
Loaded manipulated_data_no_outleirs from ../pickles/manipulated_data_no_outleirs.p
kl
Loaded player_shots from ../pickles/player_shots.pkl
Loaded df_after_EDA from ../pickles/df_after_EDA.pkl
Loaded teamstats from ../pickles/teamstats.pkl
Loaded df_combined from ../pickles/df_combined.pkl
```

descriptive statistics

	gameID	leagueID	season	date	homeTeamID	awayTeamID	homeGoals	awayGoals	
0	81	1	2015	2015-08-08 15:45:00	89	82	1	0	
1	82	1	2015	2015-08-08 18:00:00	73	71	0	1	
2	83	1	2015	2015-08-08 18:00:00	72	90	2	2	
3	84	1	2015	2015-08-08 18:00:00	75	77	4	2	
4	85	1	2015	2015-08-08 18:00:00	79	78	1	3	
...	...	...	...	...	...	...	...	...	
12675	16131	5	2020	2021-05-23 19:00:00	168	166	1	2	
12676	16132	5	2020	2021-05-23 19:00:00	177	176	1	2	
12677	16133	5	2020	2021-05-23 19:00:00	163	235	2	0	
12678	16134	5	2020	2021-05-23 19:00:00	175	181	0	1	
12679	16135	5	2020	2021-05-23 19:00:00	225	179	1	1	

12680 rows × 47 columns

```
Defaulting to user installation because normal site-packages is not writeable
Requirement already satisfied: Autoviz in /home/leoadmin/.local/lib/python3.8/site
-packages (0.1.902)
Requirement already satisfied: emoji in /home/leoadmin/.local/lib/python3.8/site-p
ackages (from Autoviz) (2.14.1)
Requirement already satisfied: fsspec>=0.8.3 in /home/leoadmin/.local/lib/python3.
8/site-packages (from Autoviz) (2025.3.0)
Requirement already satisfied: holoviews<=1.14.9 in /home/leoadmin/.local/lib/pyth
on3.8/site-packages (from Autoviz) (1.14.9)
Requirement already satisfied: hvplot~=0.7.3 in /home/leoadmin/.local/lib/python3.
8/site-packages (from Autoviz) (0.7.3)
Requirement already satisfied: matplotlib<=3.7.4 in /home/leoadmin/.local/lib/pyth
on3.8/site-packages (from Autoviz) (3.7.4)
Requirement already satisfied: nltk in /home/leoadmin/.local/lib/python3.8/site-pa
ckages (from Autoviz) (3.9.1)
Requirement already satisfied: numpy<1.24 in /home/leoadmin/.local/lib/python3.8/s
ite-packages (from Autoviz) (1.23.5)
Requirement already satisfied: pandas-dq>=1.29 in /home/leoadmin/.local/lib/python
3.8/site-packages (from Autoviz) (1.29)
Requirement already satisfied: pandas<2.0 in /home/leoadmin/.local/lib/python3.8/s
ite-packages (from Autoviz) (1.5.3)
Requirement already satisfied: panel~=0.14.4 in /home/leoadmin/.local/lib/python3.
8/site-packages (from Autoviz) (0.14.4)
Requirement already satisfied: param==1.13.0 in /home/leoadmin/.local/lib/python3.
8/site-packages (from Autoviz) (1.13.0)
Requirement already satisfied: pyamg in /home/leoadmin/.local/lib/python3.8/site-p
ackages (from Autoviz) (5.1.0)
Requirement already satisfied: scikit-learn in /home/leoadmin/.local/lib/python3.
8/site-packages (from Autoviz) (1.3.2)
Requirement already satisfied: seaborn<=0.12.2 in /home/leoadmin/.local/lib/python
3.8/site-packages (from Autoviz) (0.12.2)
Requirement already satisfied: statsmodels in /home/leoadmin/.local/lib/python3.8/
site-packages (from Autoviz) (0.14.1)
Requirement already satisfied: textblob in /home/leoadmin/.local/lib/python3.8/sit
e-packages (from Autoviz) (0.18.0.post0)
Requirement already satisfied: typing-extensions>=4.1.1 in /home/leoadmin/.local/l
ib/python3.8/site-packages (from Autoviz) (4.12.2)
Requirement already satisfied: wordcloud in /home/leoadmin/.local/lib/python3.8/si
te-packages (from Autoviz) (1.9.4)
Requirement already satisfied: xgboost<1.7,>=0.82 in /home/leoadmin/.local/lib/pyt
hon3.8/site-packages (from Autoviz) (1.6.2)
Requirement already satisfied: xlrd in /home/leoadmin/.local/lib/python3.8/site-pa
ckages (from Autoviz) (2.0.1)
Requirement already satisfied: pyviz-comms>=0.7.4 in /home/leoadmin/.local/lib/pyt
hon3.8/site-packages (from holoviews<=1.14.9->Autoviz) (3.0.4)
Requirement already satisfied: colorcet in /home/leoadmin/.local/lib/python3.8/sit
e-packages (from holoviews<=1.14.9->Autoviz) (3.1.0)
Requirement already satisfied: packaging in /home/leoadmin/.local/lib/python3.8/si
te-packages (from holoviews<=1.14.9->Autoviz) (24.2)
Requirement already satisfied: bokeh>=1.0.0 in /home/leoadmin/.local/lib/python3.
8/site-packages (from hvplot~=0.7.3->Autoviz) (2.4.3)
Requirement already satisfied: contourpy>=1.0.1 in /home/leoadmin/.local/lib/pytho
n3.8/site-packages (from matplotlib<=3.7.4->Autoviz) (1.1.1)
Requirement already satisfied: cycler>=0.10 in /home/leoadmin/.local/lib/python3.
8/site-packages (from matplotlib<=3.7.4->Autoviz) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /home/leoadmin/.local/lib/pyth
on3.8/site-packages (from matplotlib<=3.7.4->Autoviz) (4.56.0)
Requirement already satisfied: kiwisolver>=1.0.1 in /home/leoadmin/.local/lib/pyth
on3.8/site-packages (from matplotlib<=3.7.4->Autoviz) (1.4.7)
```

```
Requirement already satisfied: pillow>=6.2.0 in /home/leoadmin/.local/lib/python3.
8/site-packages (from matplotlib<=3.7.4->Autoviz) (10.4.0)
Requirement already satisfied: pyparsing>=2.3.1 in /home/leoadmin/.local/lib/pytho
n3.8/site-packages (from matplotlib<=3.7.4->Autoviz) (3.1.4)
Requirement already satisfied: python-dateutil>=2.7 in /home/leoadmin/.local/lib/p
ython3.8/site-packages (from matplotlib<=3.7.4->Autoviz) (2.9.0.post0)
Requirement already satisfied: importlib-resources>=3.2.0 in /home/leoadmin/.loca
l/lib/python3.8/site-packages (from matplotlib<=3.7.4->Autoviz) (6.4.5)
Requirement already satisfied: pytz>=2020.1 in /home/leoadmin/.local/lib/python3.
8/site-packages (from pandas<2.0->Autoviz) (2025.1)
Requirement already satisfied: markdown in /home/leoadmin/.local/lib/python3.8/sit
e-packages (from panel~=0.14.4->Autoviz) (3.7)
Requirement already satisfied: requests in /home/leoadmin/.local/lib/python3.8/sit
e-packages (from panel~=0.14.4->Autoviz) (2.32.3)
Requirement already satisfied: tqdm>=4.48.0 in /home/leoadmin/.local/lib/python3.
8/site-packages (from panel~=0.14.4->Autoviz) (4.67.1)
Requirement already satisfied: pyct>=0.4.4 in /home/leoadmin/.local/lib/python3.8/
site-packages (from panel~=0.14.4->Autoviz) (0.5.0)
Requirement already satisfied: bleach in /home/leoadmin/.local/lib/python3.8/site-
packages (from panel~=0.14.4->Autoviz) (6.1.0)
Requirement already satisfied: setuptools>=42 in /home/leoadmin/.local/lib/python
3.8/site-packages (from panel~=0.14.4->Autoviz) (75.3.0)
Requirement already satisfied: scipy>=1.5.0 in /home/leoadmin/.local/lib/python3.
8/site-packages (from scikit-learn->Autoviz) (1.10.1)
Requirement already satisfied: joblib>=1.1.1 in /home/leoadmin/.local/lib/python3.
8/site-packages (from scikit-learn->Autoviz) (1.4.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in /home/leoadmin/.local/lib/p
ython3.8/site-packages (from scikit-learn->Autoviz) (3.5.0)
Requirement already satisfied: click in /usr/lib/python3/dist-packages (from nltk-
>Autoviz) (7.0)
Requirement already satisfied: regex>=2021.8.3 in /home/leoadmin/.local/lib/python
3.8/site-packages (from nltk->Autoviz) (2024.11.6)
Requirement already satisfied: patsy>=0.5.4 in /home/leoadmin/.local/lib/python3.
8/site-packages (from statsmodels->Autoviz) (1.0.1)
Requirement already satisfied: Jinja2>=2.9 in /home/leoadmin/.local/lib/python3.8/
site-packages (from bokeh>=1.0.0->hvplot~=0.7.3->Autoviz) (3.1.6)
Requirement already satisfied: PyYAML>=3.10 in /usr/lib/python3/dist-packages (fro
m bokeh>=1.0.0->hvplot~=0.7.3->Autoviz) (5.3.1)
Requirement already satisfied: tornado>=5.1 in /home/leoadmin/.local/lib/python3.
8/site-packages (from bokeh>=1.0.0->hvplot~=0.7.3->Autoviz) (6.4.2)
Requirement already satisfied: zipp>=3.1.0 in /home/leoadmin/.local/lib/python3.8/
site-packages (from importlib-resources>=3.2.0->matplotlib<=3.7.4->Autoviz) (3.20.
2)
Requirement already satisfied: six>=1.5 in /usr/lib/python3/dist-packages (from py
thon-dateutil>=2.7->matplotlib<=3.7.4->Autoviz) (1.14.0)
Requirement already satisfied: webencodings in /home/leoadmin/.local/lib/python3.
8/site-packages (from bleach->panel~=0.14.4->Autoviz) (0.5.1)
Requirement already satisfied: importlib-metadata>=4.4 in /home/leoadmin/.local/li
b/python3.8/site-packages (from markdown->panel~=0.14.4->Autoviz) (8.5.0)
Requirement already satisfied: charset-normalizer<4,>=2 in /home/leoadmin/.local/l
ib/python3.8/site-packages (from requests->panel~=0.14.4->Autoviz) (3.4.1)
Requirement already satisfied: idna<4,>=2.5 in /usr/lib/python3/dist-packages (fro
m requests->panel~=0.14.4->Autoviz) (2.8)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/lib/python3/dist-package
s (from requests->panel~=0.14.4->Autoviz) (1.25.8)
Requirement already satisfied: certifi>=2017.4.17 in /home/leoadmin/.local/lib/pyt
hon3.8/site-packages (from requests->panel~=0.14.4->Autoviz) (2025.1.31)
Requirement already satisfied: MarkupSafe>=2.0 in /home/leoadmin/.local/lib/python
3.8/site-packages (from Jinja2>=2.9->bokeh>=1.0.0->hvplot~=0.7.3->Autoviz) (2.1.5)
```

```
Shape of your Data Set loaded: (12680, 47)
################################################################################
#####
####################### C L A S S I F Y I N G  V A R I A B L E S  ##############
#####
################################################################################
#####
Classifying variables in data set...
    Number of Numeric Columns =  16
    Number of Integer-Categorical Columns =  27
    Number of String-Categorical Columns =  1
    Number of Factor-Categorical Columns =  0
    Number of String-Boolean Columns =  0
    Number of Numeric-Boolean Columns =  0
    Number of Discrete String Columns =  0
    Number of NLP String Columns =  1
    Number of Date Time Columns =  1
    Number of ID Columns =  1
    Number of Columns to Delete =  0
    47 Predictors classified...
        1 variable(s) removed since they were ID or low-information variables
        List of variables removed: ['gameID']
16 numeric variables in data exceeds limit, taking top 30 variables
    List of variables selected: ['home_xGoals', 'home_ppda', 'home_total_xAssist
s', 'home_total_xGoalsChain', 'home_total_xGoalsBuildup', 'home_total_blocked_shot
s', 'home_total_saved_shots', 'away_xGoals', 'away_ppda', 'away_total_xAssists',
'away_total_xGoalsChain', 'away_total_xGoalsBuildup', 'away_total_blocked_shots',
'home_yellowCards', 'away_yellowCards', 'away_total_saved_shots']
   Total columns > 30, too numerous to print.
To fix these data quality issues in the dataset, import FixDQ from autoviz...
    All variables classified into correct types.
```

	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ
gameID	int64	0.000000	100	81.000000	16135.000000	Possible ID colum before modelin
leagueID	int64	0.000000	0	1.000000	5.000000	N
season	int64	0.000000	0	2014.000000	2020.000000	Possible date-time transform before mo
date	object	0.000000	53			N
homeTeamID	int64	0.000000	1	71.000000	262.000000	Column has 55 c greater than upper (256.00) or low lower bound(8.0 them or remove
awayTeamID	int64	0.000000	1	71.000000	262.000000	Column has 55 c greater than upper (256.00) or low lower bound(8.0 them or remove
homeGoals	int64	0.000000	0	0.000000	10.000000	Column has 981 c greater than upper (3.50) or lower thar bound(-0.50). Cap t remove
awayGoals	int64	0.000000	0	0.000000	9.000000	Column has 44 c greater than upper (5.00) or lower thar bound(-3.00). Cap t remove
homeGoalsHalfTime	int64	0.000000	0	0.000000	6.000000	Column has 403 c greater than upper (2.50) or lower thar bound(-1.50). Cap t remove
awayGoalsHalfTime	int64	0.000000	0	0.000000	5.000000	Column has 225 c greater than upper (2.50) or lower thar bound(-1.50). Cap t remove
home_xGoals	float64	0.000000	NA	0.000000	6.630490	Column has 276 c greater than upper (3.76) or lower thar bound(-0.92). Cap t remove
home_shots	int64	0.000000	0	0.000000	47.000000	Column has 167 c greater than upper (27.50) or lower thar bound(-0.50). Cap t remove
home_shotsOnTarget	int64	0.000000	0	0.000000	18.000000	Column has 351 c greater than upper (10.50) or lower thar bound(-1.50). Cap t

	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ
						remove them., Colum high correlatic ['home_total_saved_ Consider dropping
home_deep	int64	0.000000	0	0.000000	42.000000	Column has 227 c greater than upper (18.00) or lower thar bound(-6.00). Cap t remove
home_ppda	float64	0.000000	NA	1.897400	97.333300	Column has 540 c greater than upper (21.31) or lower thar bound(-1.88). Cap t remove
home_fouls	int64	0.000000	0	0.000000	33.000000	Column has 241 c greater than upper (22.50) or lower thar bound(2.50). Cap t remove
home_corners	int64	0.000000	0	0.000000	20.000000	Column has 143 c greater than upper (13.00) or lower thar bound(-3.00). Cap t remove
home_yellowCards	float64	0.007886	NA	0.000000	8.000000	1 missing values. them with mean, m mode, or a constan such as 123., Column outliers greater thar bound (6.00) or low lower bound(-2.0(them or remove
home_redCards	int64	0.000000	0	0.000000	3.000000	Column has 1078 c greater than upper (0.00) or lower thar bound(0.00). Cap t remove
home_total_assists	int64	0.000000	0	0.000000	8.000000	Column has 20 c greater than upper (5.00) or lower thar bound(-3.00). Cap t remove them., Colum high correlatic ['homeGoals']. Co dropping one o
home_total_xAssists	float64	0.000000	NA	0.000000	5.512622	Column has 346 c greater than upper (2.76) or lower thar bound(-0.79). Cap t remove them., Colum high correlatic ['home_xGoals']. Co dropping one o

	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ
home_total_key_passes	int64	0.000000	0	0.000000	38.000000	Column has 123 c greater than upper (22.00) or lower thar bound(-2.00). Cap t remove them., Colum high correlatic ['home_shots']. Cc dropping one o
home_total_xGoalsChain	float64	0.000000	NA	0.000000	32.011994	Column has 523 c greater than upper (10.99) or lower thar bound(-3.97). Cap t remove them., Colum high correlatic ['home_total_xAs Consider dropping
home_total_xGoalsBuildup	float64	0.000000	NA	0.000000	24.437683	Column has 664 c greater than upper (6.79) or lower thar bound(-2.88). Cap t remove them., Colum high correlatic ['home_total_xGoalsC Consider dropping
home_total_yellow_cards	int64	0.000000	0	0.000000	8.000000	Column has 20 c greater than upper (6.00) or lower thar bound(-2.00). Cap t remove them., Colum high correlatic ['home_yellowC Consider dropping
home_total_red_cards	int64	0.000000	0	0.000000	3.000000	Column has 1064 c greater than upper (0.00) or lower thar bound(0.00). Cap t remove them., Colum high correlatic ['home_redCards']. Cc dropping one o
home_total_blocked_shots	float64	0.023659	NA	0.000000	19.000000	3 missing values. them with mean, n mode, or a constan such as 123., Colu 230 outliers great upper bound (9.50) o than lower bound(Cap them or remove
home_total_saved_shots	float64	0.023659	NA	0.000000	17.000000	3 missing values. them with mean, n mode, or a constan such as 123., Colu 532 outliers great

	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ
						upper bound (7.00) o than lower bound(Cap them or remove
away_xGoals	float64	0.000000	NA	0.000000	6.186960	Column has 312 c greater than upper (3.10) or lower thar bound(-0.91). Cap t remove
away_shots	int64	0.000000	0	0.000000	39.000000	Column has 161 c greater than upper (23.00) or lower thar bound(-1.00). Cap t remove
away_shotsOnTarget	int64	0.000000	0	0.000000	15.000000	Column has 233 c greater than upper (9.50) or lower thar bound(-2.50). Cap t remove them., Colum high correlatic ['away_total_saved_ Consider dropping
away_deep	int64	0.000000	0	0.000000	28.000000	Column has 423 c greater than upper (13.00) or lower thar bound(-3.00). Cap t remove
away_ppda	float64	0.000000	NA	2.122000	152.000000	Column has 606 c greater than upper (24.25) or lower thar bound(-2.58). Cap t remove
away_fouls	int64	0.000000	0	0.000000	32.000000	Column has 81 c greater than upper (25.00) or lower thar bound(1.00). Cap t remove
away_corners	int64	0.000000	0	0.000000	19.000000	Column has 283 c greater than upper (10.50) or lower thar bound(-1.50). Cap t remove
away_yellowCards	float64	0.000000	NA	0.000000	9.000000	Column has 43 c greater than upper (6.00) or lower thar bound(-2.00). Cap t remove
away_redCards	int64	0.000000	0	0.000000	3.000000	Column has 1396 c greater than upper (0.00) or lower thar bound(0.00). Cap t remove

	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ
away_total_assists	int64	0.000000	0	0.000000	8.000000	Column has 790 c greater than upper (2.50) or lower thar bound(-1.50). Cap t remove them., Colum high correlatic ['awayGoals']. Co dropping one o
away_total_xAssists	float64	0.000000	NA	0.000000	5.463750	Column has 389 c greater than upper (2.29) or lower thar bound(-0.77). Cap t remove them., Colum high correlatic ['away_xGoals']. Co dropping one o
away_total_key_passes	int64	0.000000	0	0.000000	27.000000	Column has 169 c greater than upper (18.50) or lower thar bound(-1.50). Cap t remove them., Colum high correlatic ['away_shots']. Co dropping one o
away_total_xGoalsChain	float64	0.000000	NA	0.000000	34.587459	Column has 538 c greater than upper (9.16) or lower thar bound(-3.62). Cap t remove them., Colum high correlatic ['away_total_xAs Consider dropping
away_total_xGoalsBuildup	float64	0.000000	NA	0.000000	27.419105	Column has 718 c greater than upper (5.57) or lower thar bound(-2.50). Cap t remove them., Colum high correlatic ['away_total_xGoalsC Consider dropping
away_total_yellow_cards	int64	0.000000	0	0.000000	9.000000	Column has 26 c greater than upper (6.00) or lower thar bound(-2.00). Cap t remove them., Colum high correlatic ['away_yellowC Consider dropping
away_total_red_cards	int64	0.000000	0	0.000000	3.000000	Column has 1382 c greater than upper (0.00) or lower thar bound(0.00). Cap t remove them., Colum

	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ
						high correlatic ['away_redCards']. Co dropping one o
away_total_blocked_shots	float64	0.063091	NA	0.000000	16.000000	8 missing values. them with mean, n mode, or a constan such as 123., Colu 187 outliers great upper bound (8.50) o than lower bound(Cap them or remove
away_total_saved_shots	float64	0.063091	NA	0.000000	13.000000	8 missing values. them with mean, n mode, or a constan such as 123., Column outliers greater thar bound (8.50) or low lower bound(-3.5(them or remove
gameresult	object	0.000000	0			N

```
Number of All Scatter Plots = 136
```

```
[nltk_data] Downloading collection 'popular'
[nltk_data]    |
[nltk_data]    | Downloading package cmudict to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package cmudict is already up-to-date!
[nltk_data]    | Downloading package gazetteers to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package gazetteers is already up-to-date!
[nltk_data]    | Downloading package genesis to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package genesis is already up-to-date!
[nltk_data]    | Downloading package gutenberg to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package gutenberg is already up-to-date!
[nltk_data]    | Downloading package inaugural to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package inaugural is already up-to-date!
[nltk_data]    | Downloading package movie_reviews to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package movie_reviews is already up-to-date!
[nltk_data]    | Downloading package names to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package names is already up-to-date!
[nltk_data]    | Downloading package shakespeare to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package shakespeare is already up-to-date!
[nltk_data]    | Downloading package stopwords to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package stopwords is already up-to-date!
[nltk_data]    | Downloading package treebank to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package treebank is already up-to-date!
[nltk_data]    | Downloading package twitter_samples to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package twitter_samples is already up-to-date!
[nltk_data]    | Downloading package omw to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package omw is already up-to-date!
[nltk_data]    | Downloading package omw-1.4 to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package omw-1.4 is already up-to-date!
[nltk_data]    | Downloading package wordnet to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package wordnet is already up-to-date!
[nltk_data]    | Downloading package wordnet2021 to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package wordnet2021 is already up-to-date!
[nltk_data]    | Downloading package wordnet31 to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package wordnet31 is already up-to-date!
[nltk_data]    | Downloading package wordnet_ic to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package wordnet_ic is already up-to-date!
[nltk_data]    | Downloading package words to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package words is already up-to-date!
[nltk_data]    | Downloading package maxent_ne_chunker to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package maxent_ne_chunker is already up-to-date!
[nltk_data]    | Downloading package punkt to
```

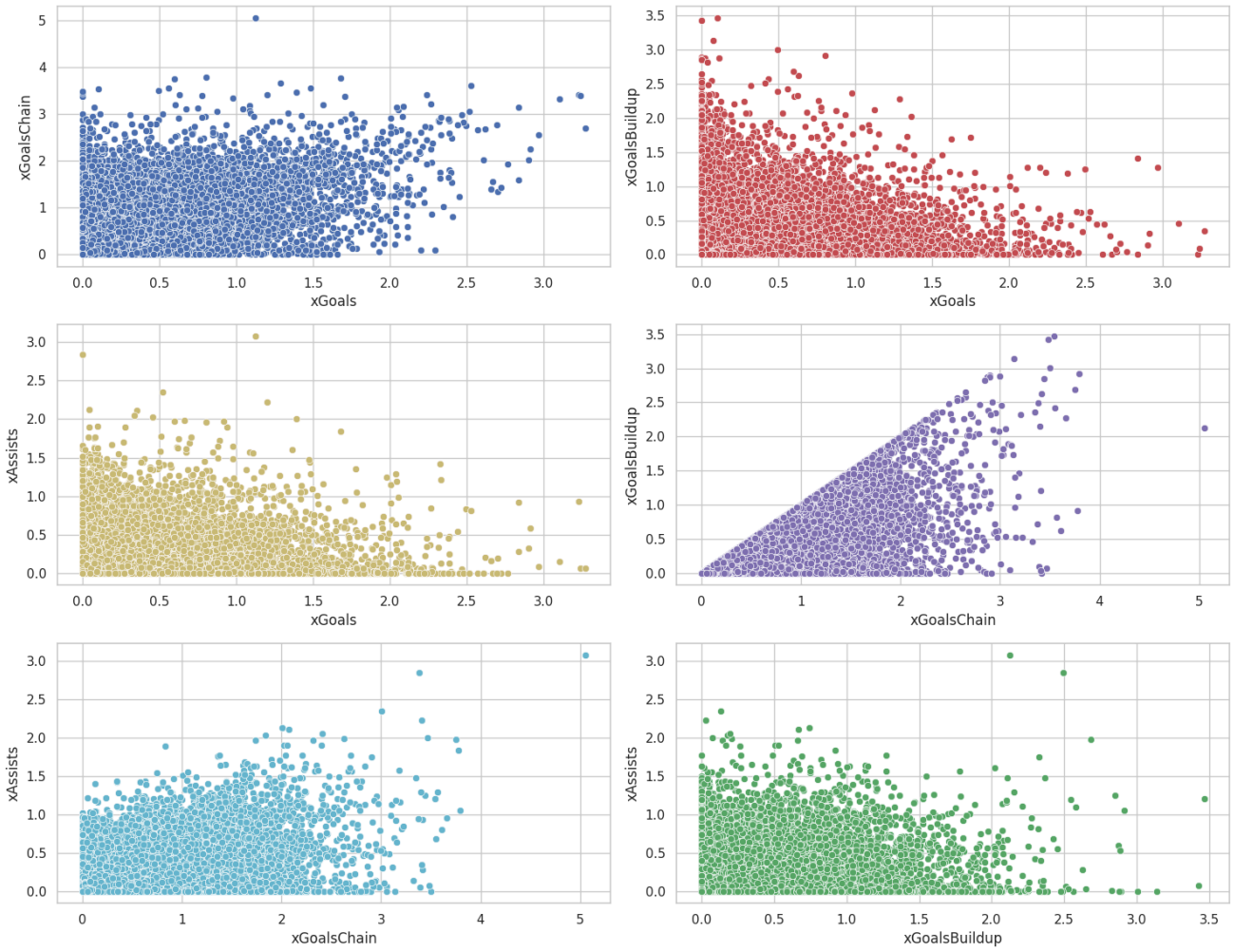
```
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package punkt is already up-to-date!
[nltk_data]    | Downloading package snowball_data to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package snowball_data is already up-to-date!
[nltk_data]    | Downloading package averaged_perceptron_tagger to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package averaged_perceptron_tagger is already up-
[nltk_data]    |       to-date!
[nltk_data]    |
[nltk_data]  Done downloading collection popular
Could not draw wordcloud plot for date
All Plots done
Time to run AutoViz = 273 seconds

 ###################### AUTO VISUALIZATION Completed ######################


=== Running AutoViz on df_appearances ===
    max_rows_analyzed is smaller than dataset shape 356513...
        randomly sampled 150000 rows from read CSV file
Shape of your Data Set loaded: (150000, 18)
################################################################################
#####
######################## C L A S S I F Y I N G  V A R I A B L E S  ##############
#####
################################################################################
#####
Classifying variables in data set...
    Number of Numeric Columns =  4
    Number of Integer-Categorical Columns =  9
    Number of String-Categorical Columns =  0
    Number of Factor-Categorical Columns =  0
    Number of String-Boolean Columns =  0
    Number of Numeric-Boolean Columns =  5
    Number of Discrete String Columns =  0
    Number of NLP String Columns =  0
    Number of Date Time Columns =  0
    Number of ID Columns =  0
    Number of Columns to Delete =  0
    18 Predictors classified...
        No variables removed since no ID or low-information variables found in dat
a set
Since Number of Rows in data 150000 exceeds maximum, randomly sampling 150000 rows
for EDA...
To fix these data quality issues in the dataset, import FixDQ from autoviz...
    All variables classified into correct types.
```

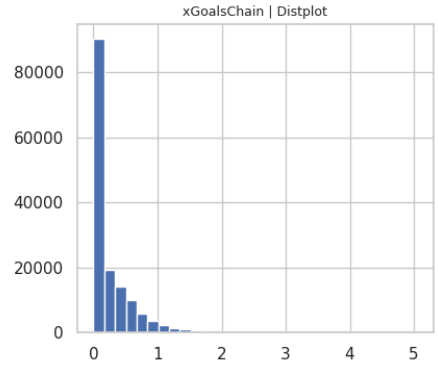
	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ Issue
gameID	int64	0.000000	8	81.000000	16135.000000	No issue
playerID	int64	0.000000	4	1.000000	9567.000000	Column has 3 outliers greater than upper bound (9547.50) or lower than lower bound(-4192.50). Cap them or remove them.
goals	int64	0.000000	0	0.000000	5.000000	Column has 12763 outliers greater than upper bound (0.00) or lower than lower bound(0.00). Cap them or remove them.
ownGoals	int64	0.000000	0	0.000000	2.000000	Column has 415 outliers greater than upper bound (0.00) or lower than lower bound(0.00). Cap them or remove them.
shots	int64	0.000000	0	0.000000	14.000000	Column has 15692 outliers greater than upper bound (2.50) or lower than lower bound(-1.50). Cap them or remove them.
xGoals	float64	0.000000	NA	0.000000	3.271321	Column has 19761 outliers greater than upper bound (0.19) or lower than lower bound(-0.12). Cap them or remove them.
xGoalsChain	float64	0.000000	NA	0.000000	5.052705	Column has 9034 outliers greater than upper bound (0.90) or lower than lower bound(-0.51). Cap them or remove them.
xGoalsBuildup	float64	0.000000	NA	0.000000	3.465713	Column has 18971 outliers greater than upper bound (0.41) or lower than lower bound(-0.25). Cap them or remove them.
assists	int64	0.000000	0	0.000000	4.000000	Column has 9165 outliers greater than upper bound (0.00) or lower than lower bound(0.00). Cap them or remove them.
keyPasses	int64	0.000000	0	0.000000	11.000000	Column has 9392 outliers greater than upper bound (2.50) or lower than lower bound(-1.50). Cap them or remove them.
xAssists	float64	0.000000	NA	0.000000	3.074537	Column has 18890 outliers greater than upper bound (0.14) or lower than lower bound(-0.08). Cap them or remove them.

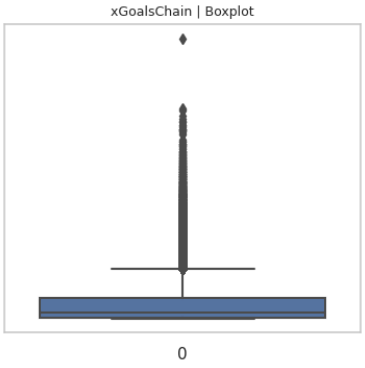
	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ Issue
positionOrder	int64	0.000000	0	1.000000	17.000000	No issue
yellowCard	int64	0.000000	0	0.000000	1.000000	No issue
redCard	int64	0.000000	0	0.000000	1.000000	No issue
time	int64	0.000000	0	1.000000	90.000000	Column has 9880 outliers greater than upper bound (138.00) or lower than lower bound(10.00). Cap them or remove them.
subOut	int64	0.000000	0	0.000000	1.000000	No issue
subIn	int64	0.000000	0	0.000000	1.000000	No issue
leagueID	int64	0.000000	0	1.000000	5.000000	No issue

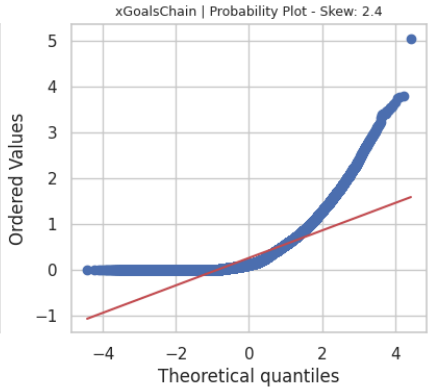
```
Number of All Scatter Plots = 10
```

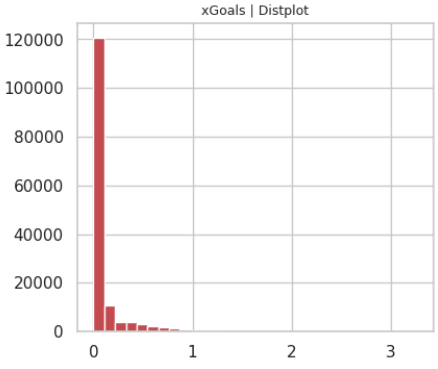
Pair-wise Scatter Plot of all Continuous Variables

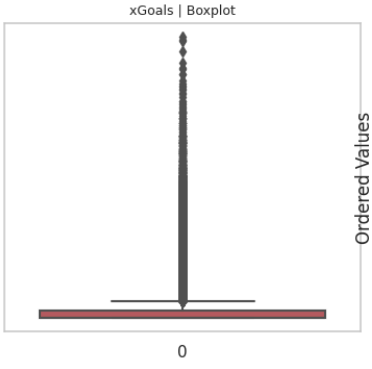


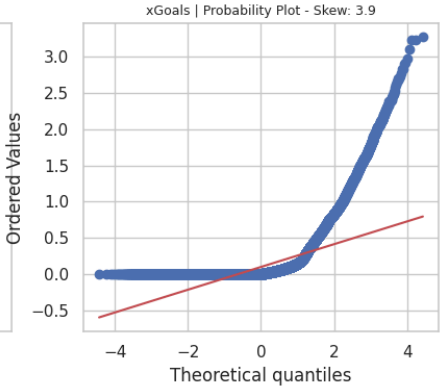

xGoalsChain | Distplot
80000
60000
40000
20000
0
0 1 2 3 4 5

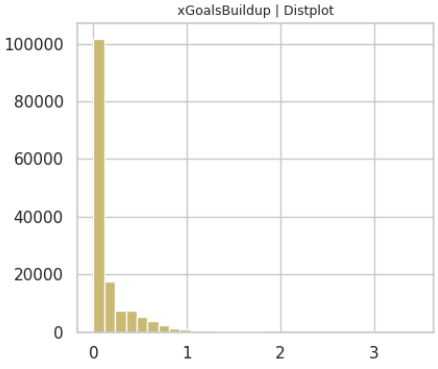

xGoalsChain | Boxplot
0

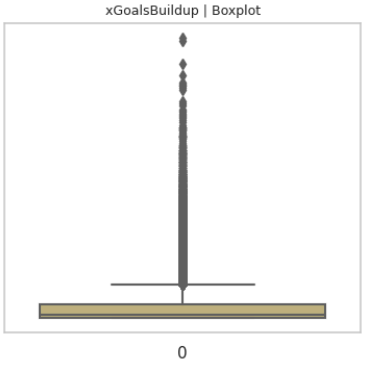

xGoalsChain | Probability Plot - Skew: 2.4
Ordered Values
5
4
3
2
1
0
−1
−4 −2 0 2 4
Theoretical quantiles

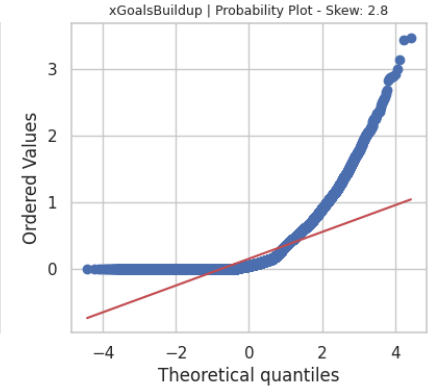

xGoals | Distplot
120000
100000
80000
60000
40000
20000
0
0 1 2 3

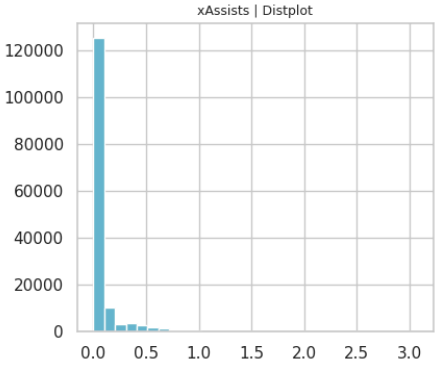

xGoals | Boxplot
0

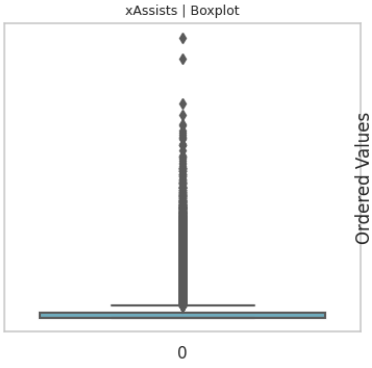

xGoals | Probability Plot - Skew: 3.9
Ordered Values
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−4 −2 0 2 4
Theoretical quantiles

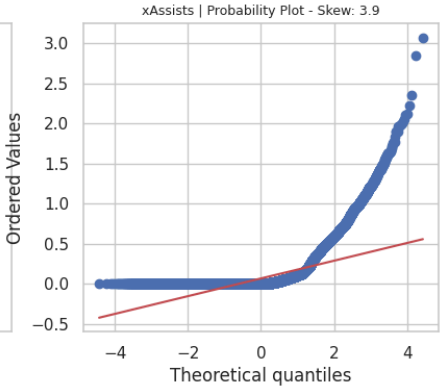

xGoalsBuildup | Distplot
100000
80000
60000
40000
20000
0
0 1 2 3

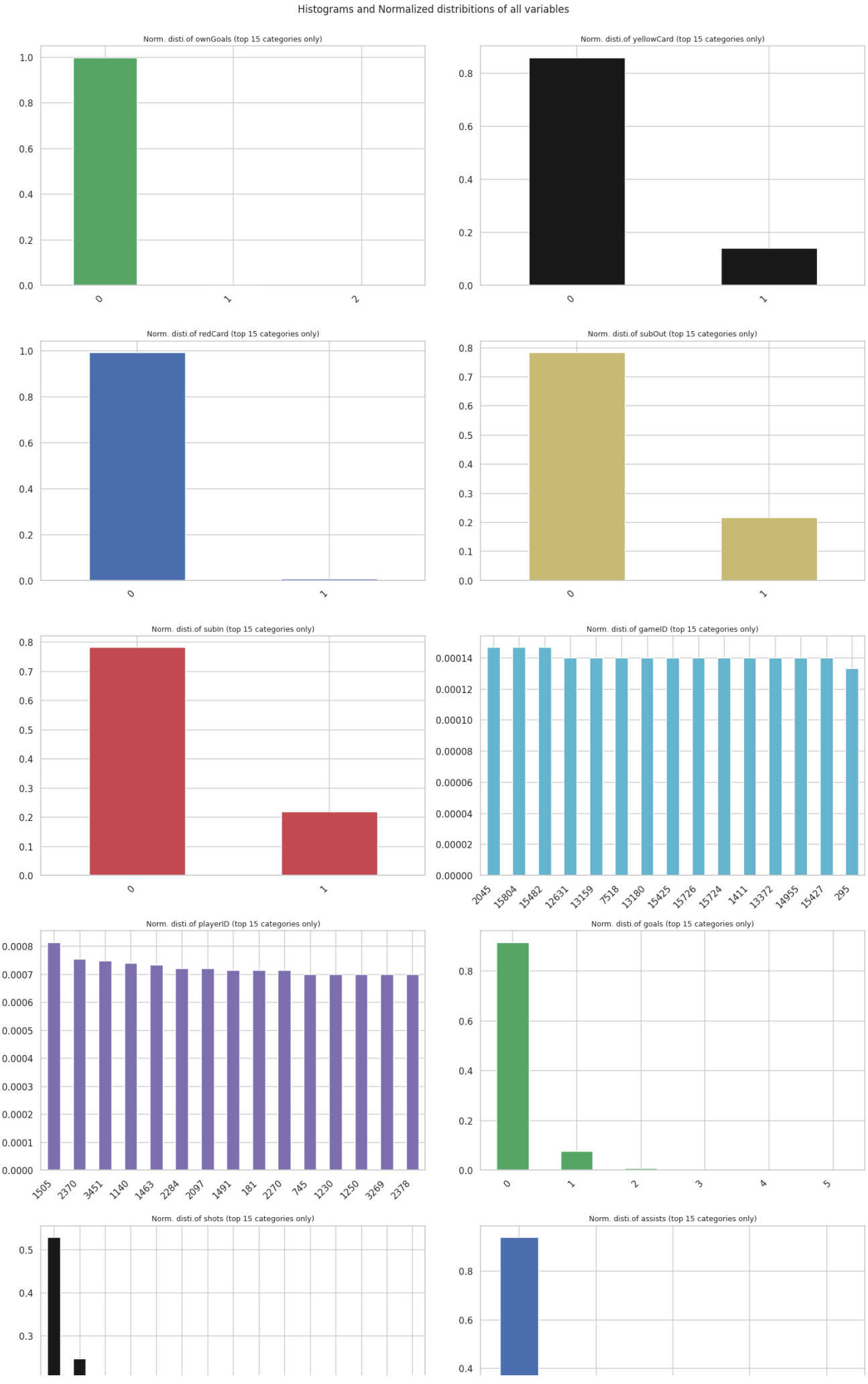

xGoalsBuildup | Boxplot
0

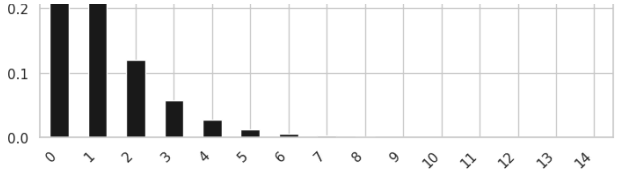

xGoalsBuildup | Probability Plot - Skew: 2.8
Ordered Values
3
2
1
0
−4 −2 0 2 4
Theoretical quantiles

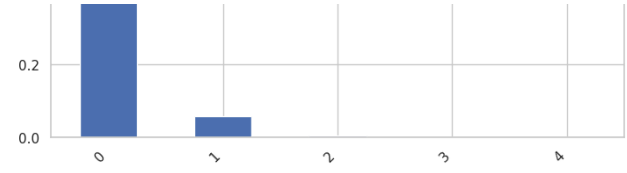

xAssists | Distplot
120000
100000
80000
60000
40000
20000
0
0.0 0.5 1.0 1.5 2.0 2.5 3.0

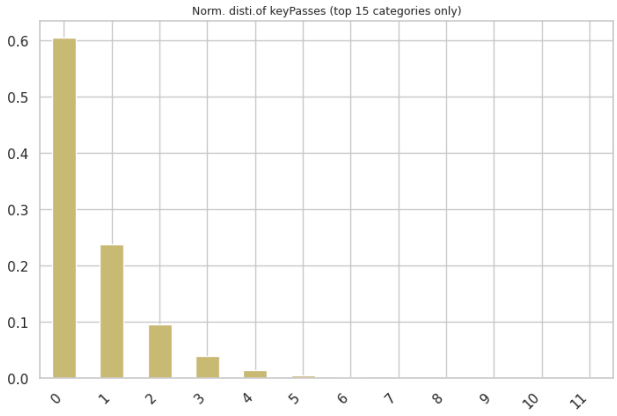

xAssists | Boxplot
0

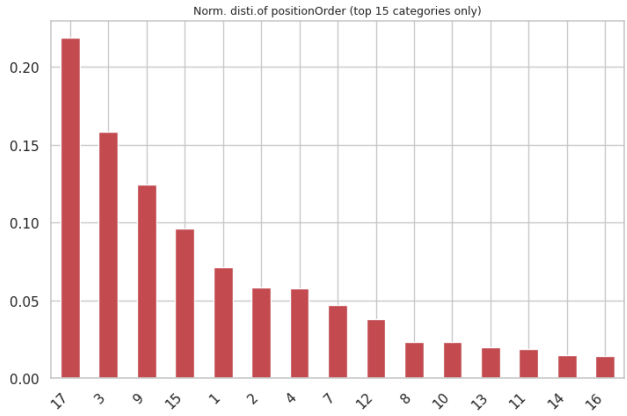

xAssists | Probability Plot - Skew: 3.9
Ordered Values
3.0
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−4 −2 0 2 4
Theoretical quantiles

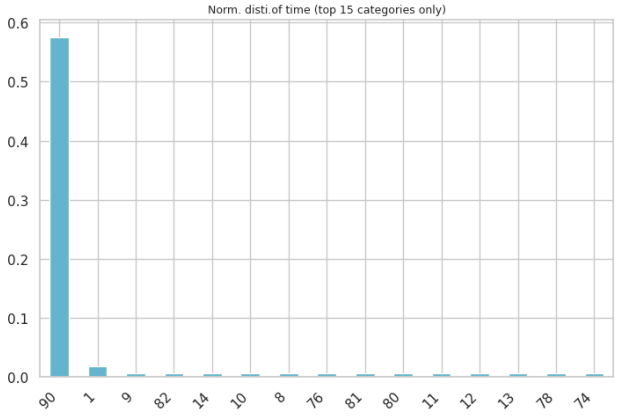

Histograms and Normalized distribitions of all variables
Norm. disti.of ownGoals (top 15 categories only)
Norm. disti.of yellowCard (top 15 categories only)
Norm. disti.of redCard (top 15 categories only)
Norm. disti.of subOut (top 15 categories only)
Norm. disti.of subIn (top 15 categories only)
Norm. disti.of gameID (top 15 categories only)
2045 15804 15482 12631 13159 7518 13180 15425 15726 15724 1411 13372 14955 15427 295
Norm. disti.of playerID (top 15 categories only)
1505 2370 3451 1140 1463 2284 2097 1491 181 2270 745 1230 1250 3269 2378
Norm. disti.of goals (top 15 categories only)
Norm. disti.of shots (top 15 categories only)
Norm. disti.of assists (top 15 categories only)

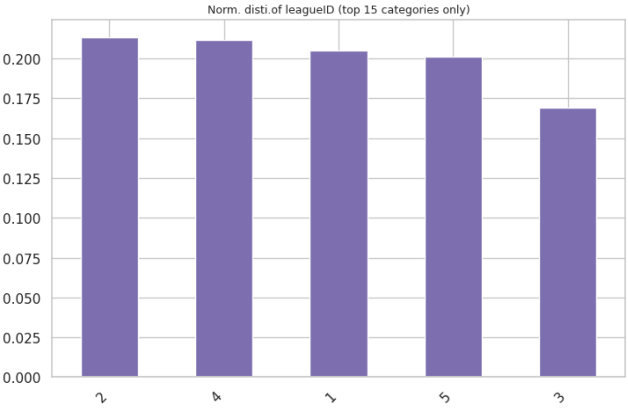


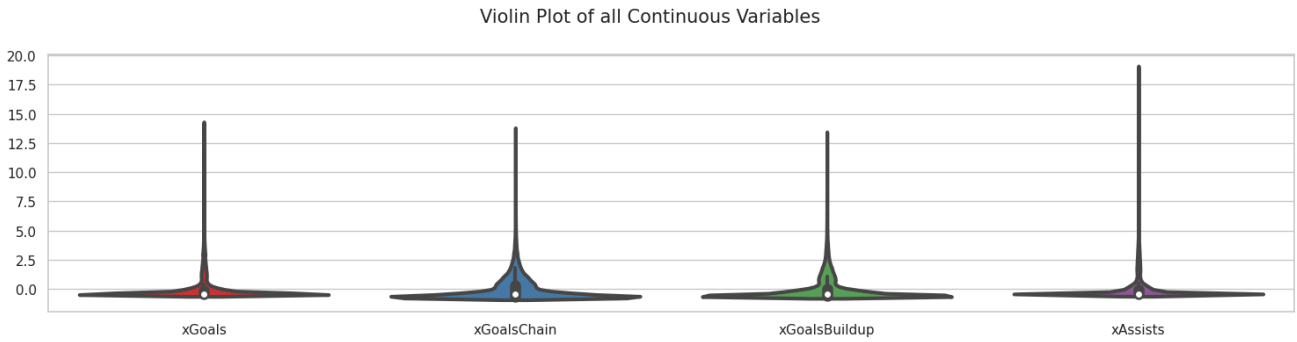



Norm. disti.of keyPasses (top 15 categories only)

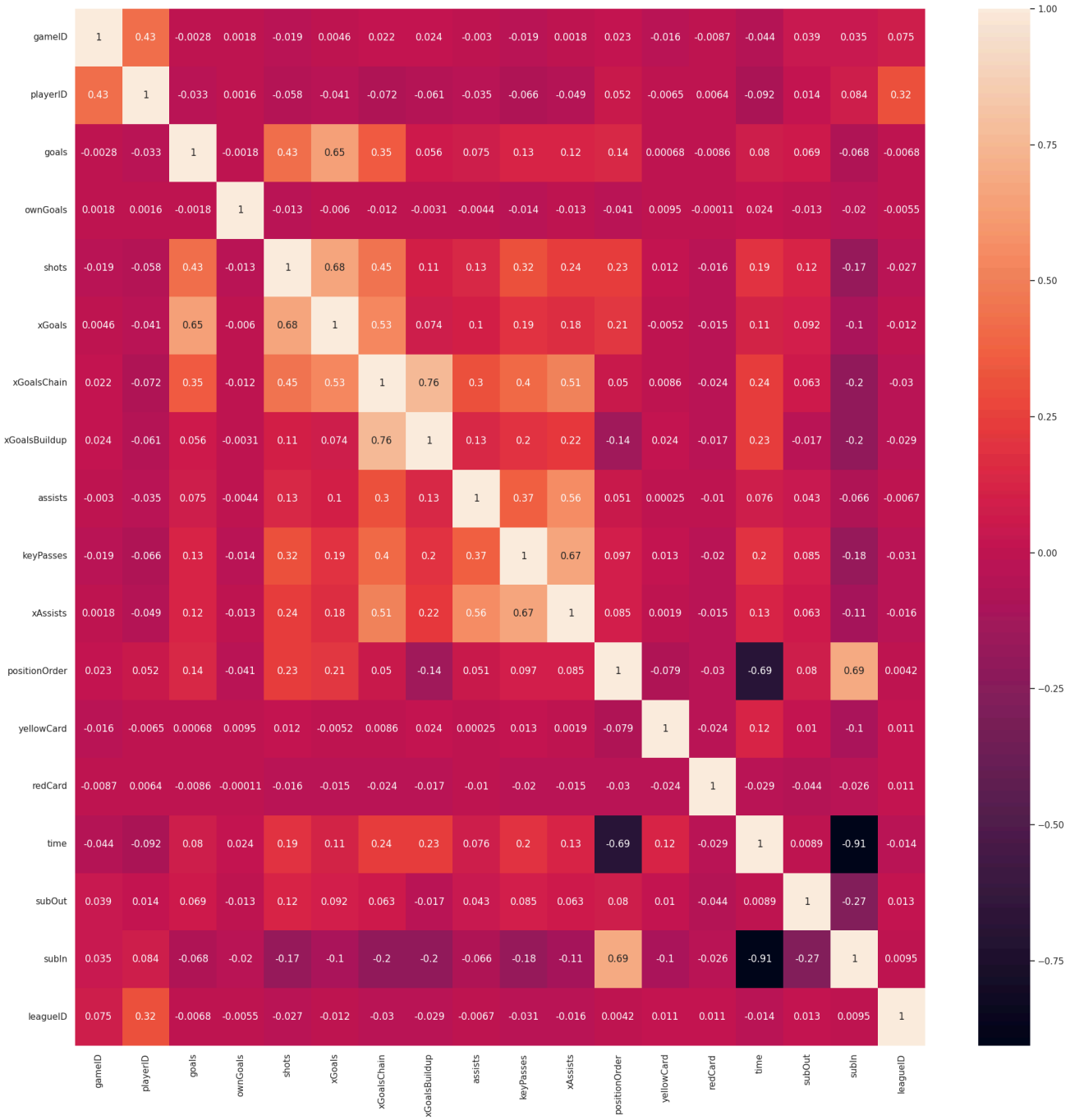

Norm. disti.of positionOrder (top 15 categories only)


Norm. disti.of time (top 15 categories only)

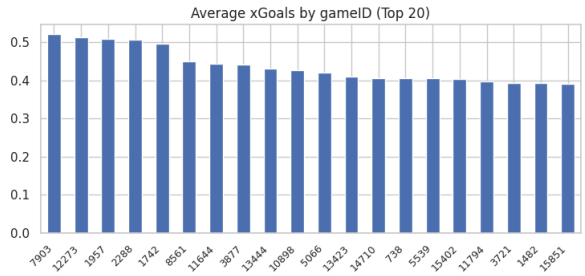

Norm. disti.of leagueID (top 15 categories only)

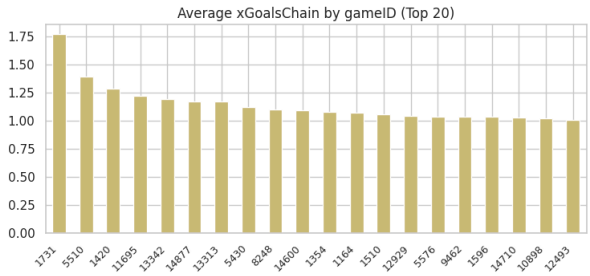

Violin Plot of all Continuous Variables
xGoals
xGoalsChain
xGoalsBuildup
xAssists

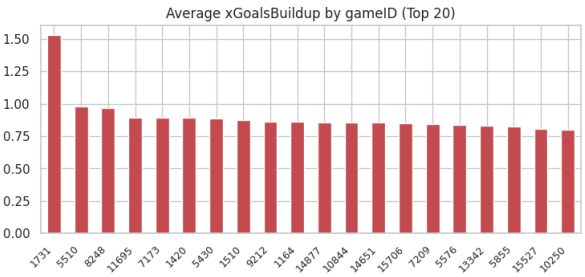
Heatmap of all Numeric Variables including target:

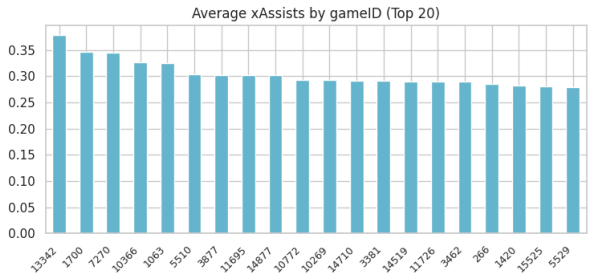


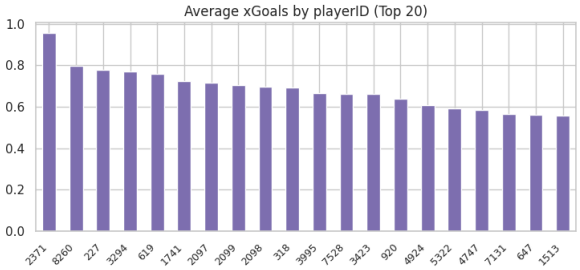
Bar plots for each Continuous by each Categorical variable

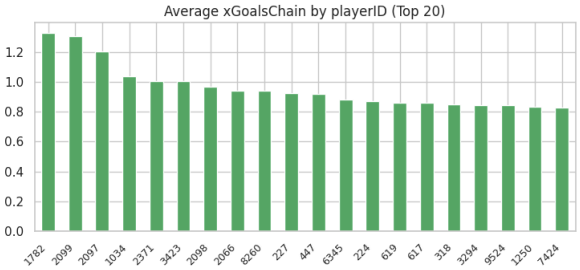


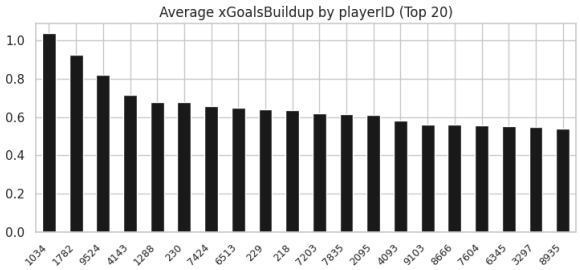


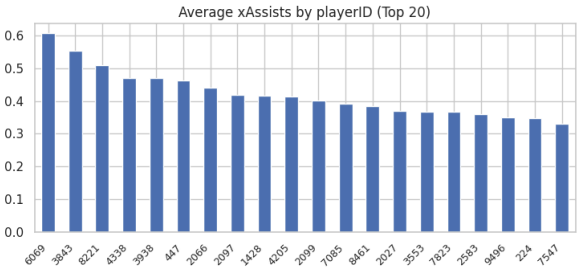


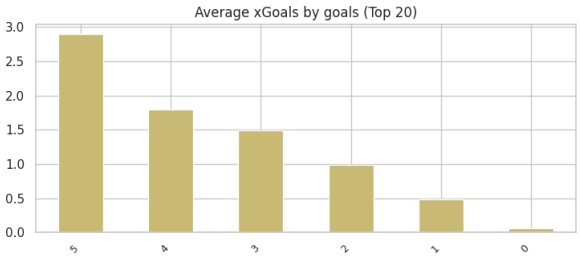


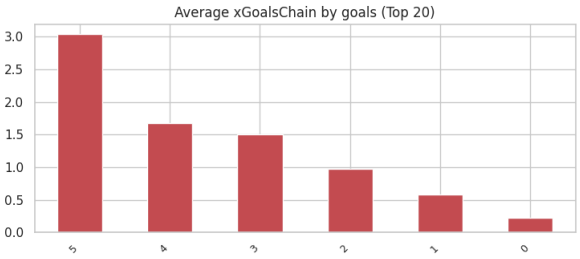

Average xGoals by playerID (Top 20)

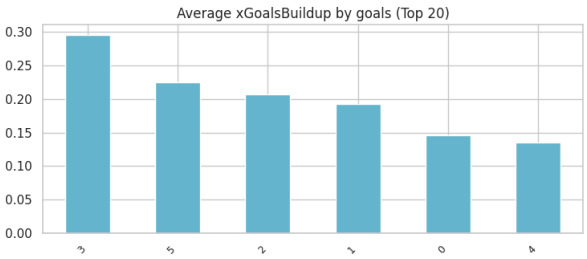

Average xGoalsChain by playerID (Top 20)

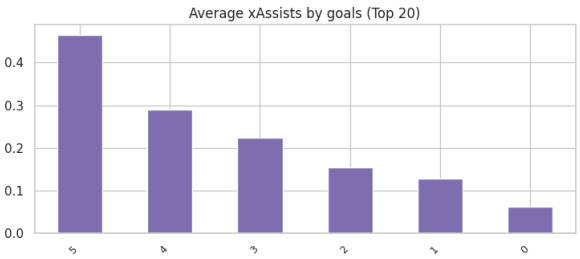

Average xGoalsBuildup by playerID (Top 20)

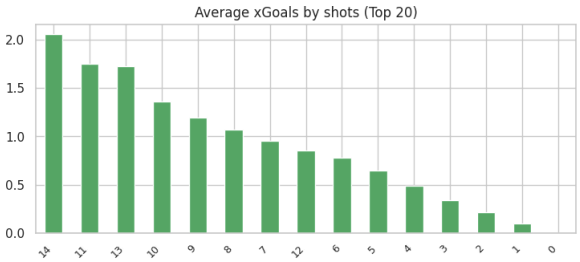

Average xAssists by playerID (Top 20)

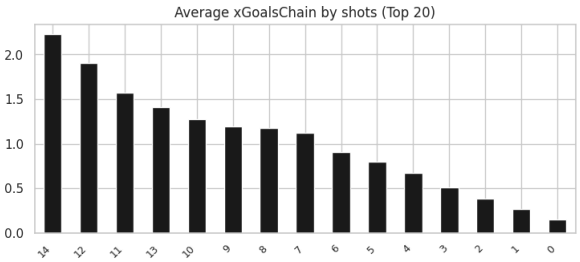

Average xGoals by goals (Top 20)

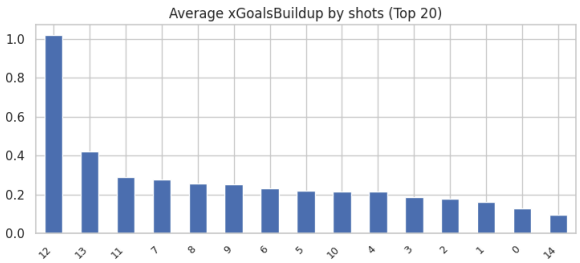

Average xGoalsChain by goals (Top 20)

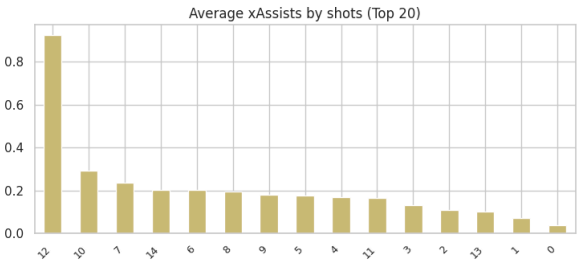

Average xGoalsBuildup by goals (Top 20)

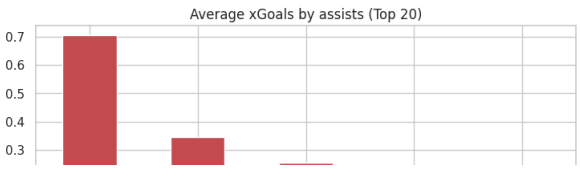

Average xAssists by goals (Top 20)

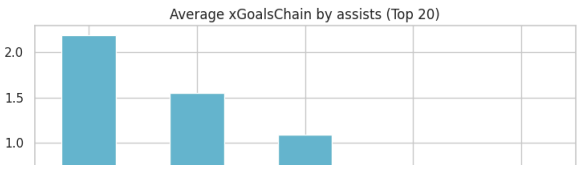

Average xGoals by shots (Top 20)

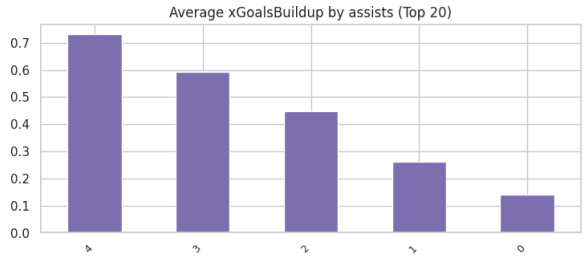

Average xGoalsChain by shots (Top 20)

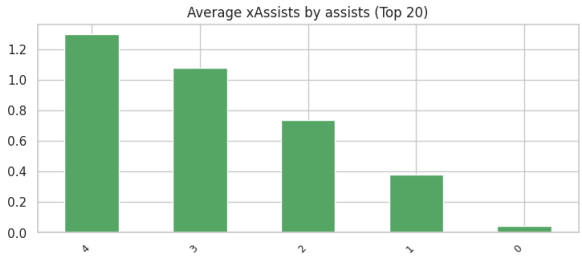

Average xGoalsBuildup by shots (Top 20)

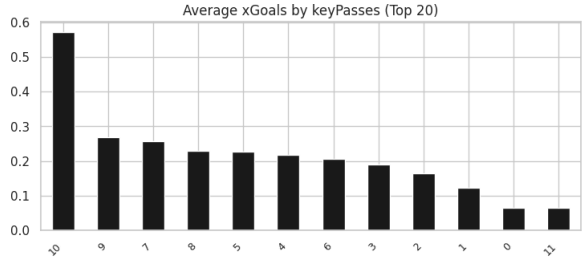

Average xAssists by shots (Top 20)

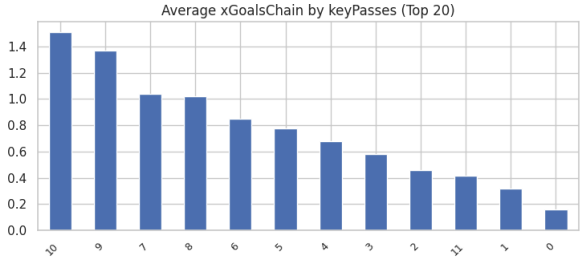

Average xGoals by assists (Top 20)

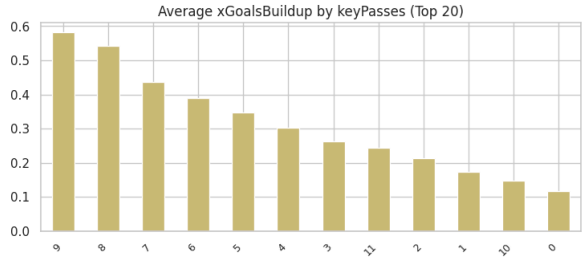

Average xGoalsChain by assists (Top 20)

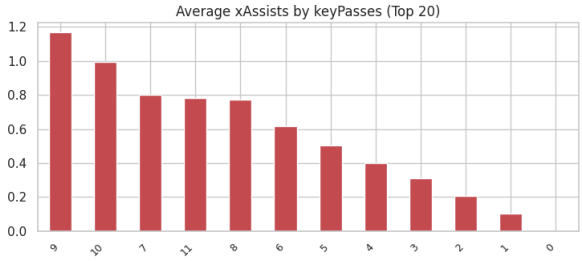

Average xGoalsBuildup by assists (Top 20)

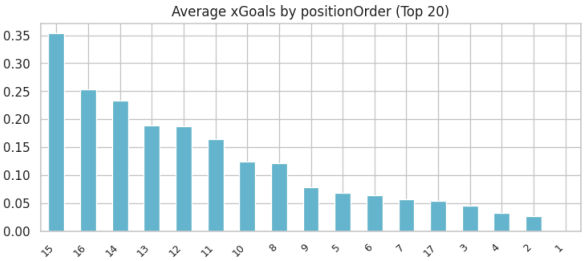

Average xAssists by assists (Top 20)

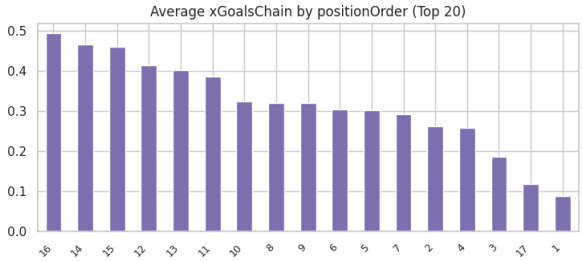

Average xGoals by keyPasses (Top 20)

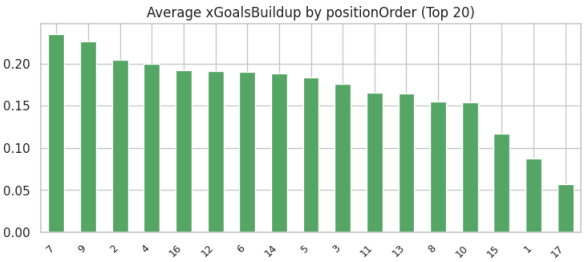

Average xGoalsChain by keyPasses (Top 20)

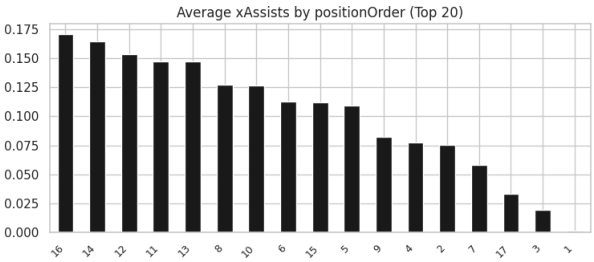

Average xGoalsBuildup by keyPasses (Top 20)

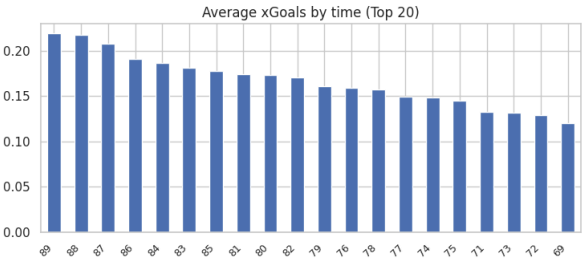

Average xAssists by keyPasses (Top 20)

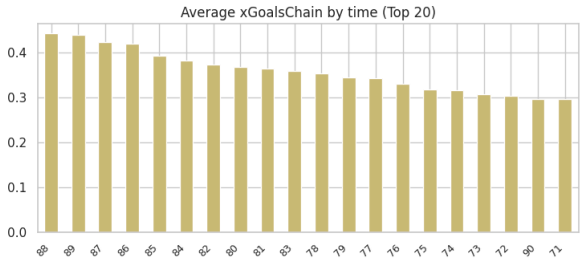

Average xGoals by positionOrder (Top 20)


Average xGoalsChain by positionOrder (Top 20)


Average xGoalsBuildup by positionOrder (Top 20)

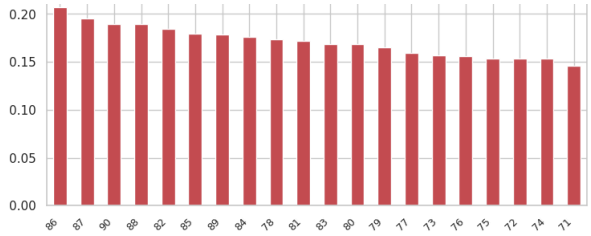

Average xAssists by positionOrder (Top 20)

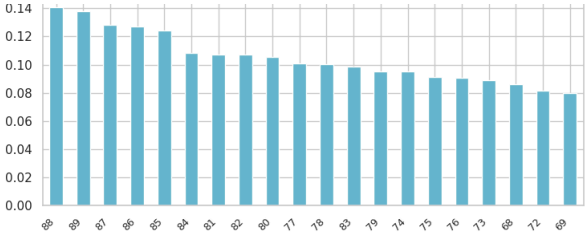

Average xGoals by time (Top 20)

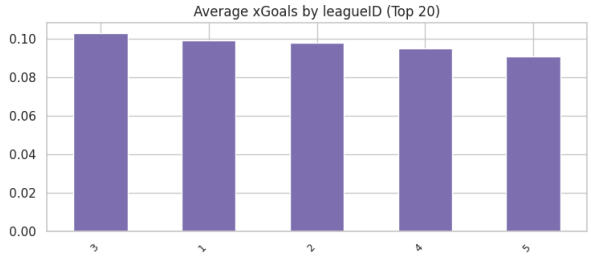

Average xGoalsChain by time (Top 20)

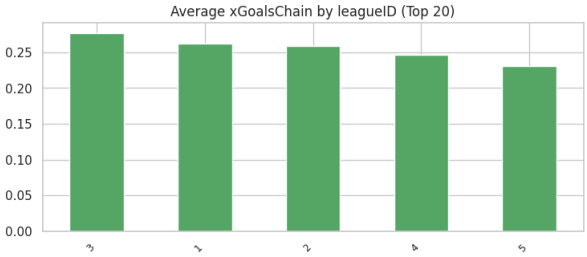

Average xGoalsBuildup by time (Top 20)

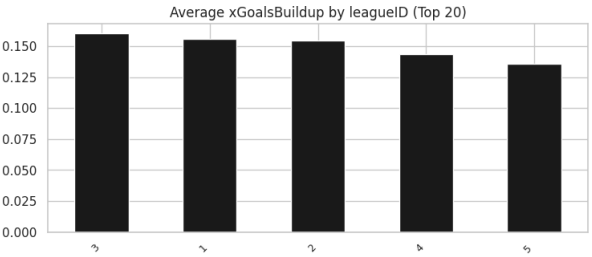

Average xAssists by time (Top 20)

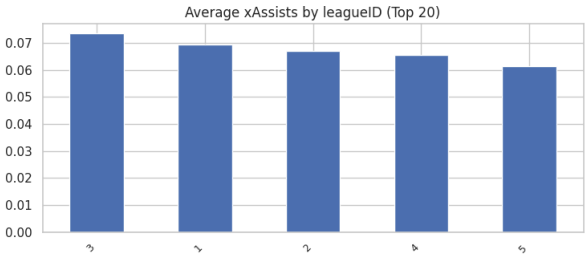


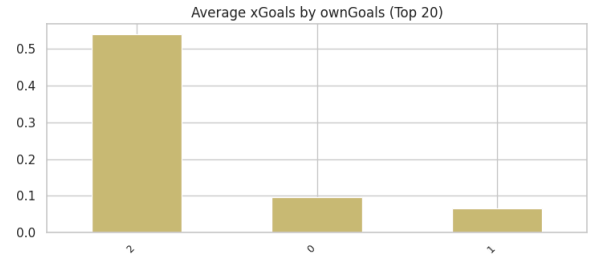


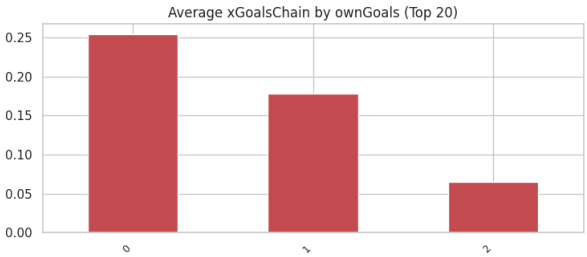

Average xGoals by leagueID (Top 20)

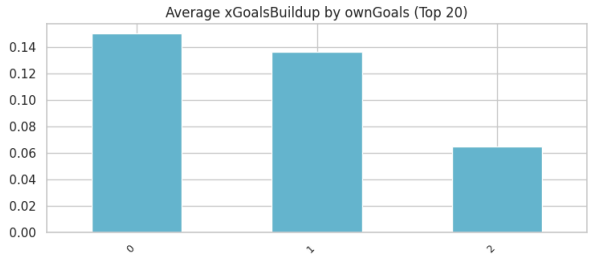

Average xGoalsChain by leagueID (Top 20)

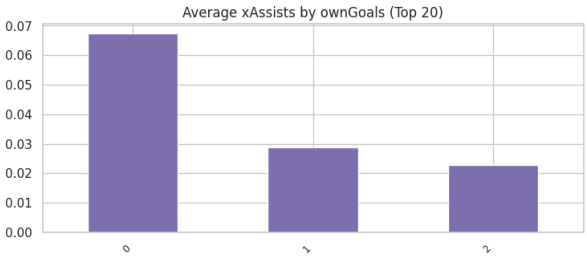

Average xGoalsBuildup by leagueID (Top 20)

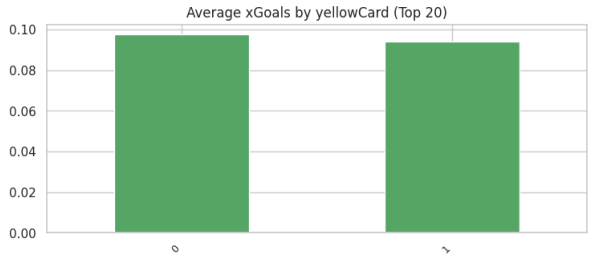

Average xAssists by leagueID (Top 20)

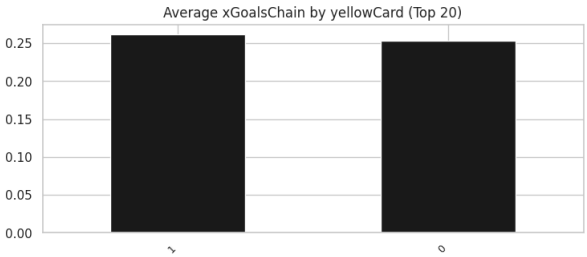

Average xGoals by ownGoals (Top 20)

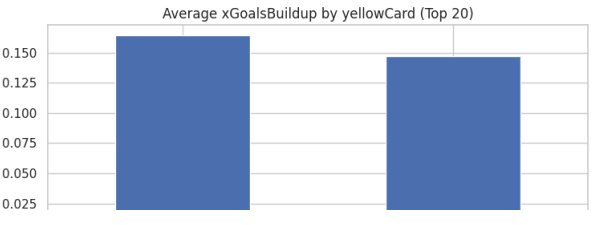

Average xGoalsChain by ownGoals (Top 20)

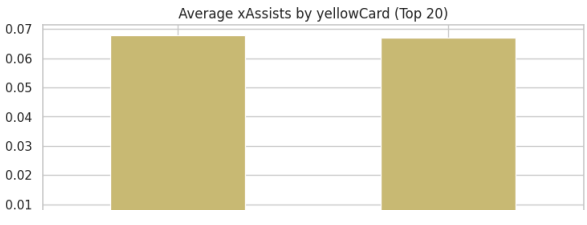

Average xGoalsBuildup by ownGoals (Top 20)


Average xAssists by ownGoals (Top 20)


Average xGoals by yellowCard (Top 20)

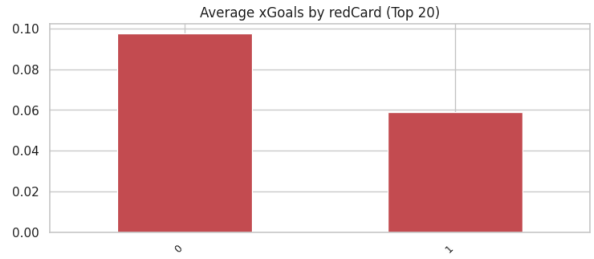

Average xGoalsChain by yellowCard (Top 20)

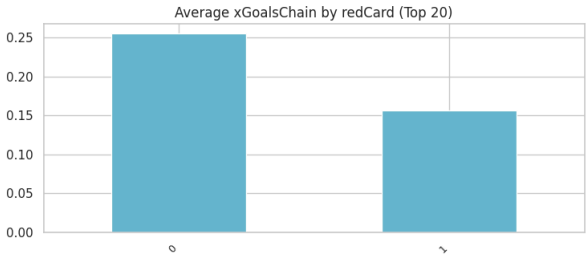

Average xGoalsBuildup by yellowCard (Top 20)

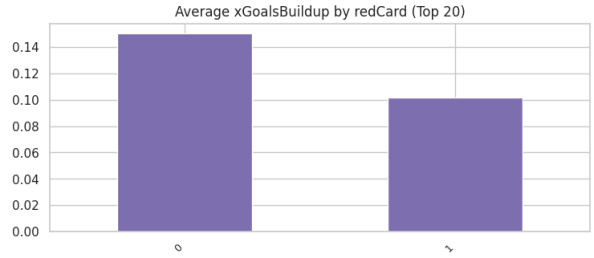

Average xAssists by yellowCard (Top 20)

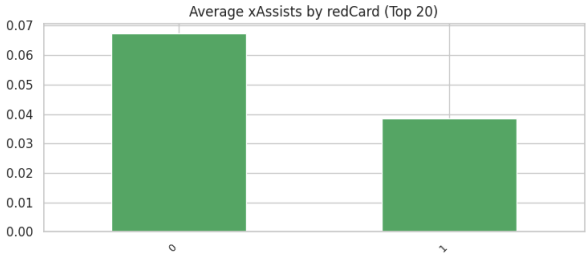

0.000
1
0

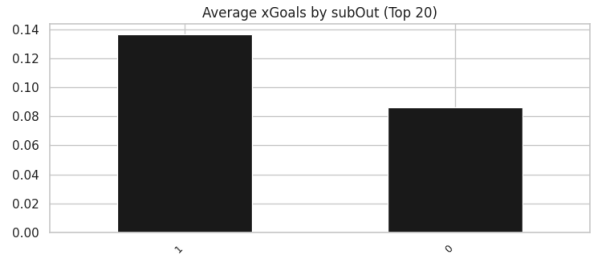

0.00
1
0

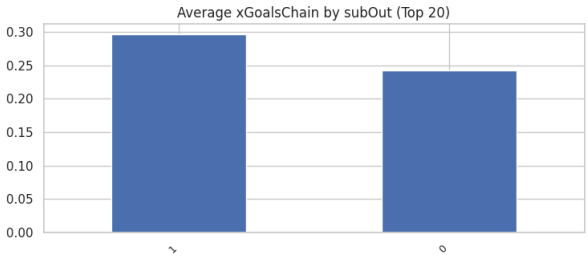

Average xGoals by redCard (Top 20)
0.10
0.08
0.06
0.04
0.02
0.00
0
1

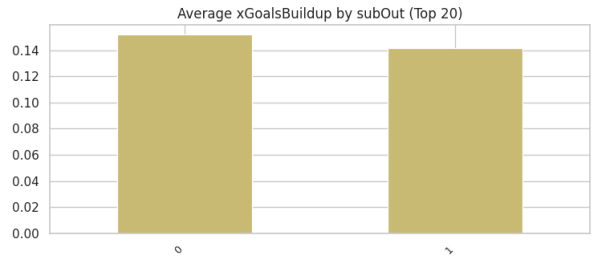

Average xGoalsChain by redCard (Top 20)
0.25
0.20
0.15
0.10
0.05
0.00
0
1

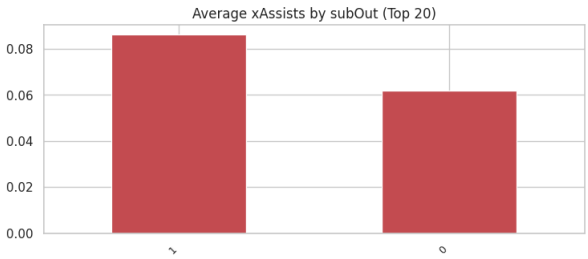

Average xGoalsBuildup by redCard (Top 20)
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
1

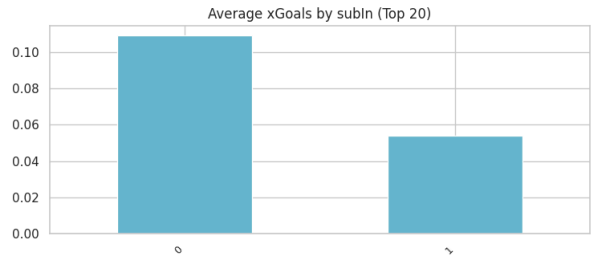

Average xAssists by redCard (Top 20)
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0.00
0
1

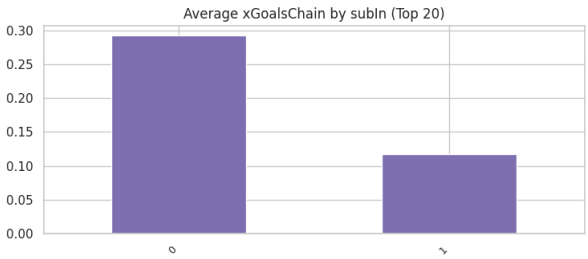

Average xGoals by subOut (Top 20)
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
1
0

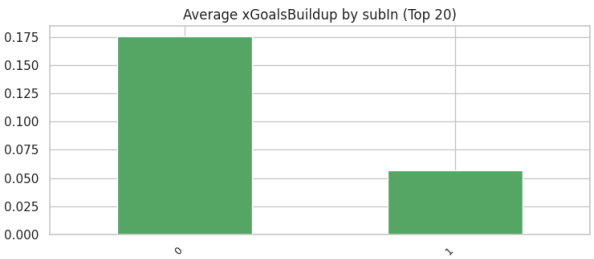

Average xGoalsChain by subOut (Top 20)
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1
0

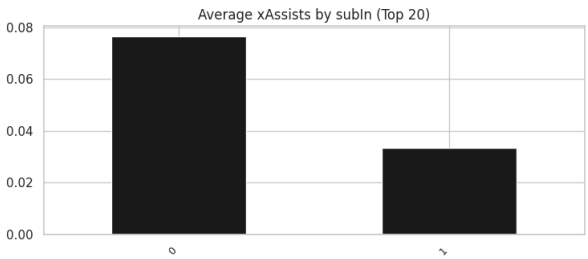

Average xGoalsBuildup by subOut (Top 20)
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
0
1


Average xAssists by subOut (Top 20)
0.08
0.06
0.04
0.02
0.00
1
0


Average xGoals by subIn (Top 20)
0.10
0.08
0.06
0.04
0.02
0.00
0
1


Average xGoalsChain by subIn (Top 20)
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0
1


Average xGoalsBuildup by subIn (Top 20)
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
0
1

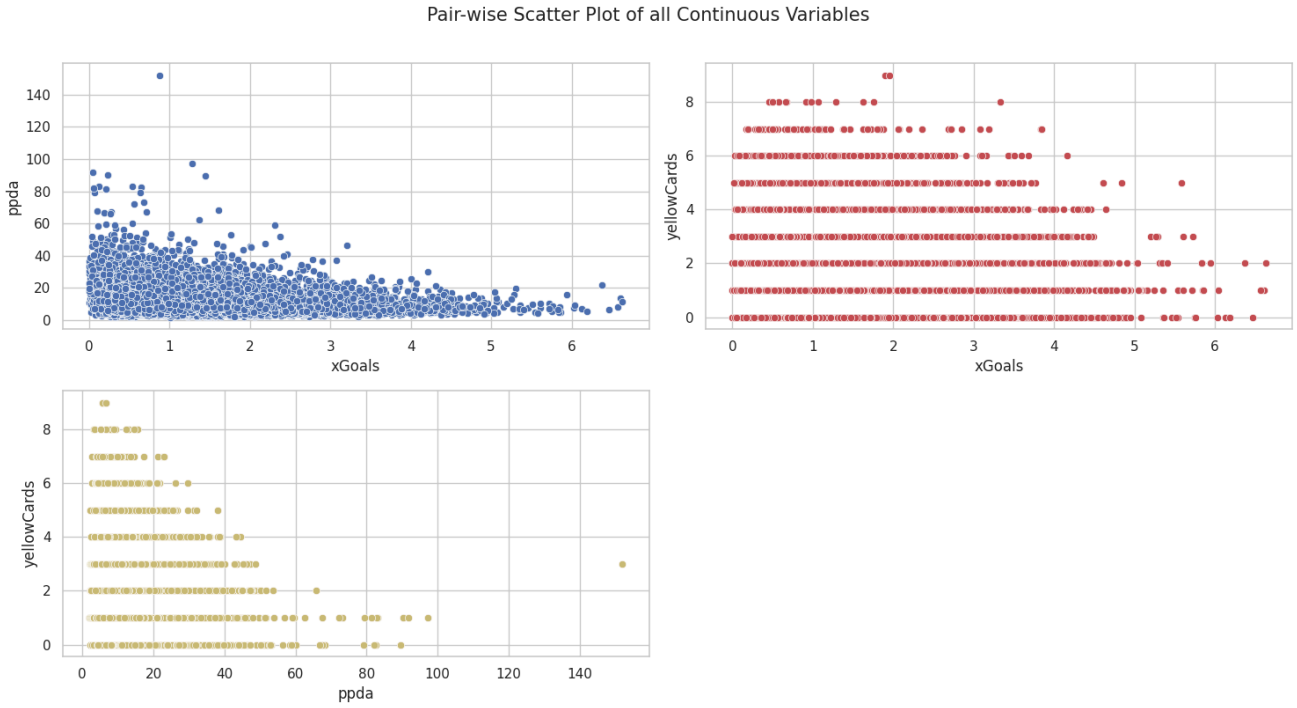

Average xAssists by subIn (Top 20)
0.08
0.06
0.04
0.02
0.00
0
1

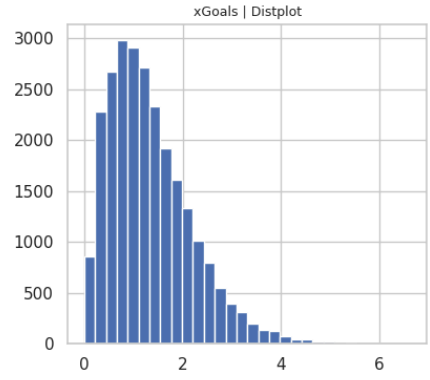
```
All Plots done
Time to run AutoViz = 23 seconds

 ###################### AUTO VISUALIZATION Completed ######################
=== Completed AutoViz on df_appearances ===

=== Running AutoViz on df_teamstats ===
Shape of your Data Set loaded: (25360, 16)
################################################################################
#####
######################## C L A S S I F Y I N G  V A R I A B L E S  ##############
#####
################################################################################
#####
Classifying variables in data set...
    Number of Numeric Columns =  3
    Number of Integer-Categorical Columns =  9
    Number of String-Categorical Columns =  1
    Number of Factor-Categorical Columns =  0
    Number of String-Boolean Columns =  1
    Number of Numeric-Boolean Columns =  0
    Number of Discrete String Columns =  0
    Number of NLP String Columns =  1
    Number of Date Time Columns =  1
    Number of ID Columns =  0
    Number of Columns to Delete =  0
    16 Predictors classified...
        No variables removed since no ID or low-information variables found in dat
a set
To fix these data quality issues in the dataset, import FixDQ from autoviz...
    All variables classified into correct types.
```

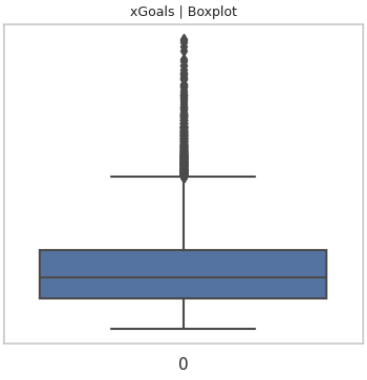
	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ Issue
gameID	int64	0.000000	50	81.000000	16135.000000	No issue
teamID	int64	0.000000	0	71.000000	262.000000	Column has 110 outliers greater than upper bound (256.00) or lower than lower bound(8.00). Cap them or remove them.
season	int64	0.000000	0	2014.000000	2020.000000	Possible date-time colum: transform before modeling step.
date	object	0.000000	26			No issue
location	object	0.000000	0			No issue
goals	int64	0.000000	0	0.000000	10.000000	Column has 156 outliers greater than upper bound (5.00) or lower than lower bound(-3.00). Cap them or remove them.
xGoals	float64	0.000000	NA	0.000000	6.630490	Column has 556 outliers greater than upper bound (3.48) or lower than lower bound(-0.97). Cap them or remove them.
shots	int64	0.000000	0	0.000000	47.000000	Column has 263 outliers greater than upper bound (26.50) or lower than lower bound(-1.50). Cap them or remove them.
shotsOnTarget	int64	0.000000	0	0.000000	18.000000	Column has 469 outliers greater than upper bound (10.50) or lower than lower bound(-1.50). Cap them or remove them.
deep	int64	0.000000	0	0.000000	42.000000	Column has 722 outliers greater than upper bound (15.50) or lower than lower bound(-4.50). Cap them or remove them.
ppda	float64	0.000000	NA	1.897400	152.000000	Column has 1177 outliers greater than upper bound (22.85) or lower than lower bound(-2.34). Cap them or remove them.
fouls	int64	0.000000	0	0.000000	33.000000	Column has 139 outliers greater than upper bound (25.00) or lower than lower bound(1.00). Cap them or remove them.
corners	int64	0.000000	0	0.000000	20.000000	Column has 178 outliers greater than upper bound (13.00) or lower than lower bound(-3.00). Cap them or remove them.

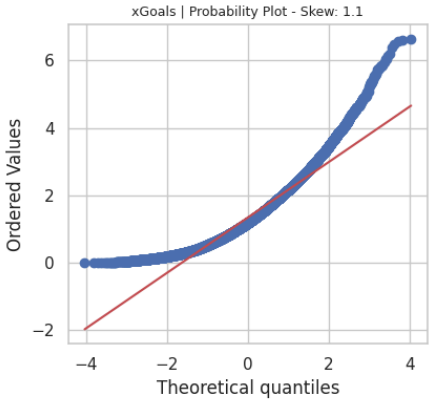
	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ Issue
yellowCards	float64	0.003943	NA	0.000000	9.000000	1 missing values. Impute them with mean, median, mode, or a constant value such as 123., Column has 80 outliers greater than upper bound (6.00) or lower than lower bound(-2.00). Cap them or remove them.
redCards	int64	0.000000	0	0.000000	3.000000	Column has 2474 outliers greater than upper bound (0.00) or lower than lower bound(0.00). Cap them or remove them.
result	object	0.000000	0			No issue

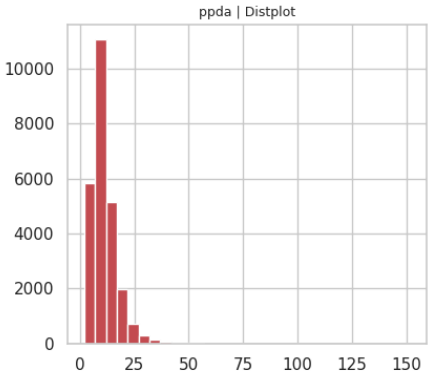
```
Number of All Scatter Plots = 6
```

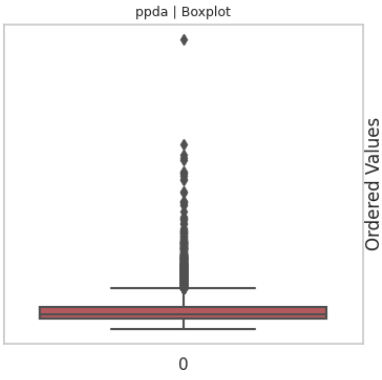


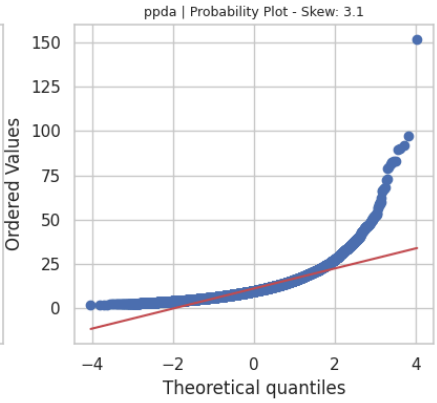

xGoals | Distplot
3000
2500
2000
1500
1000
500
0
0
2
4
6

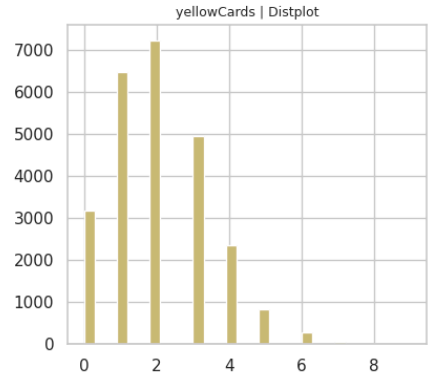

xGoals | Boxplot
0

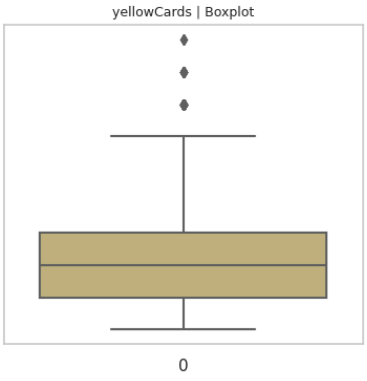

xGoals | Probability Plot - Skew: 1.1
Ordered Values
6
4
2
0
−2
−4
−2
0
2
4
Theoretical quantiles

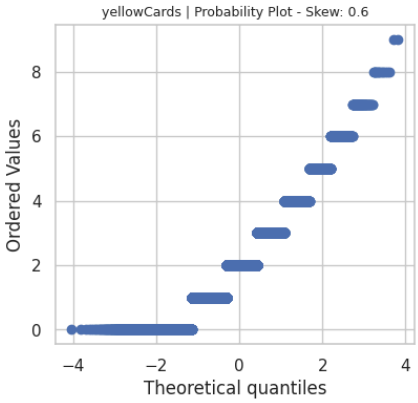

ppda | Distplot
10000
8000
6000
4000
2000
0
0
25
50
75
100
125
150

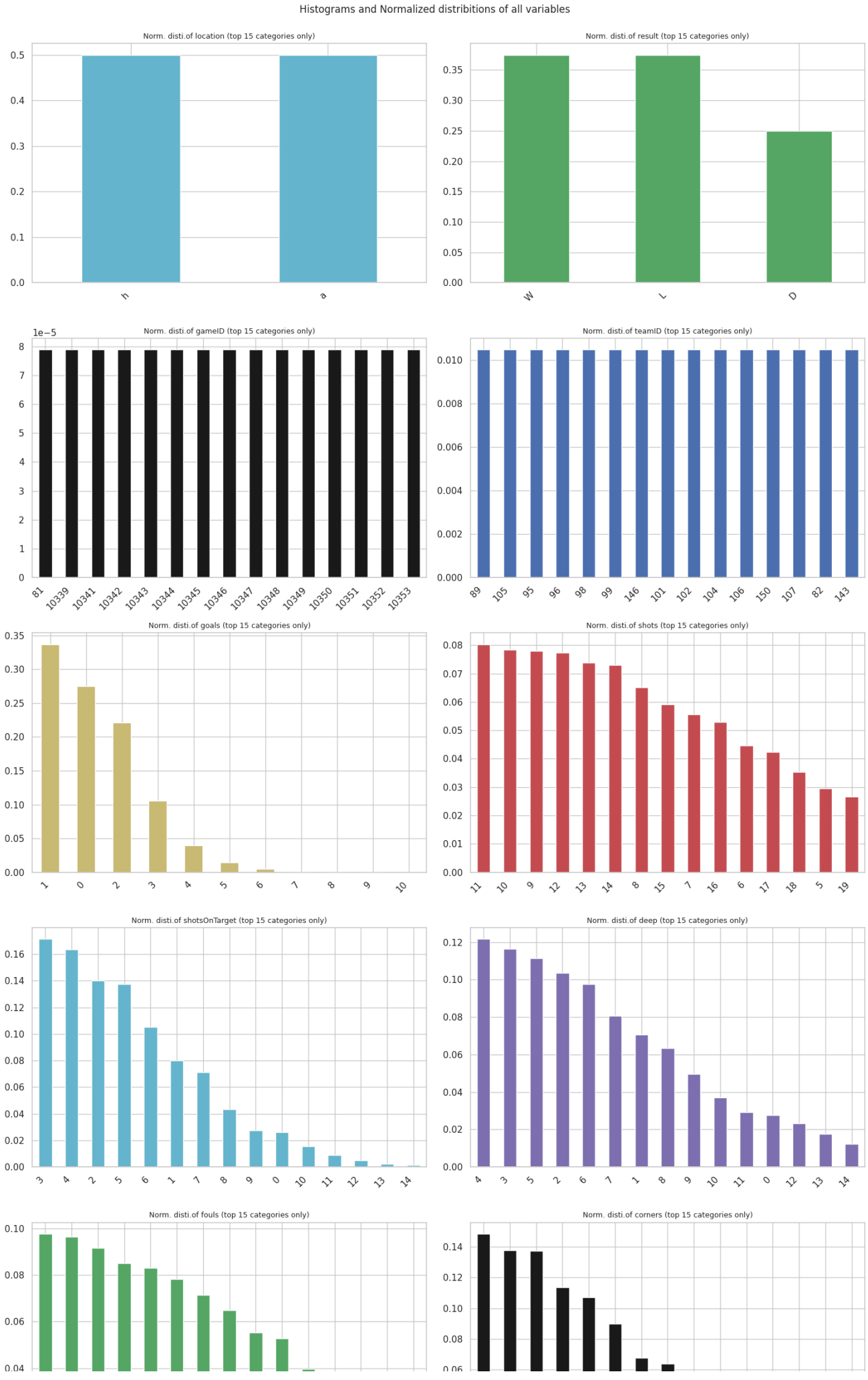

ppda | Boxplot
0

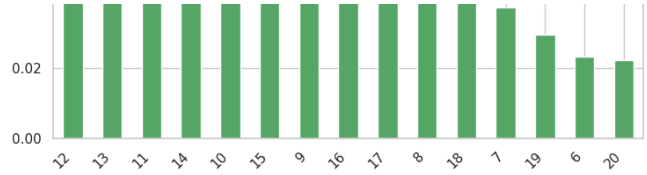

ppda | Probability Plot - Skew: 3.1
Ordered Values
150
125
100
75
50
25
0
−4
−2
0
2
4
Theoretical quantiles

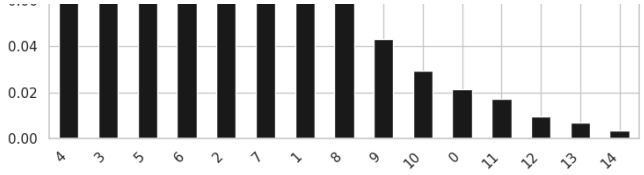

yellowCards | Distplot
7000
6000
5000
4000
3000
2000
1000
0
0
2
4
6
8

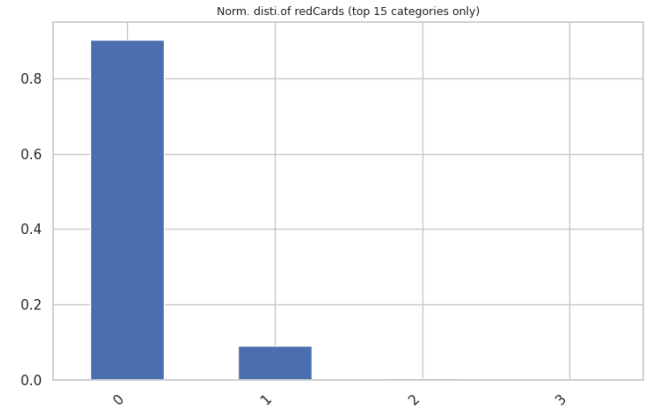

yellowCards | Boxplot
0

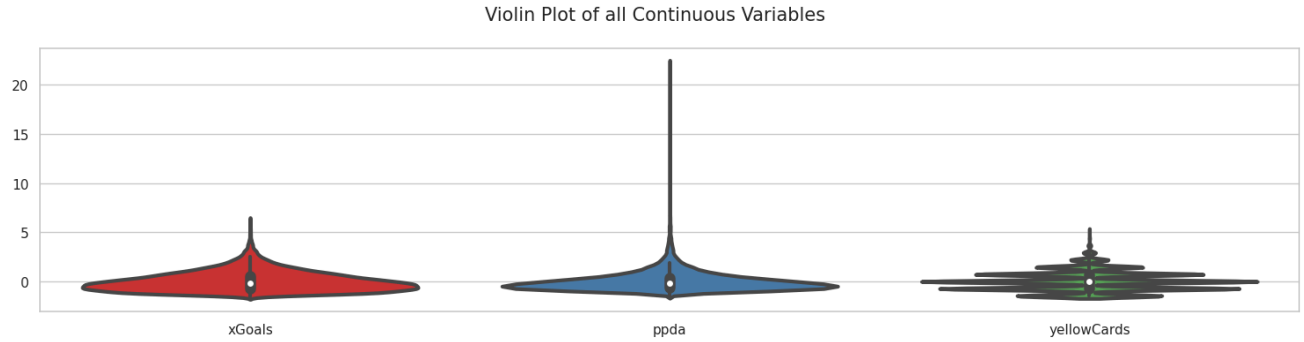

yellowCards | Probability Plot - Skew: 0.6
Ordered Values
8
6
4
2
0
−4
−2
0
2
4
Theoretical quantiles

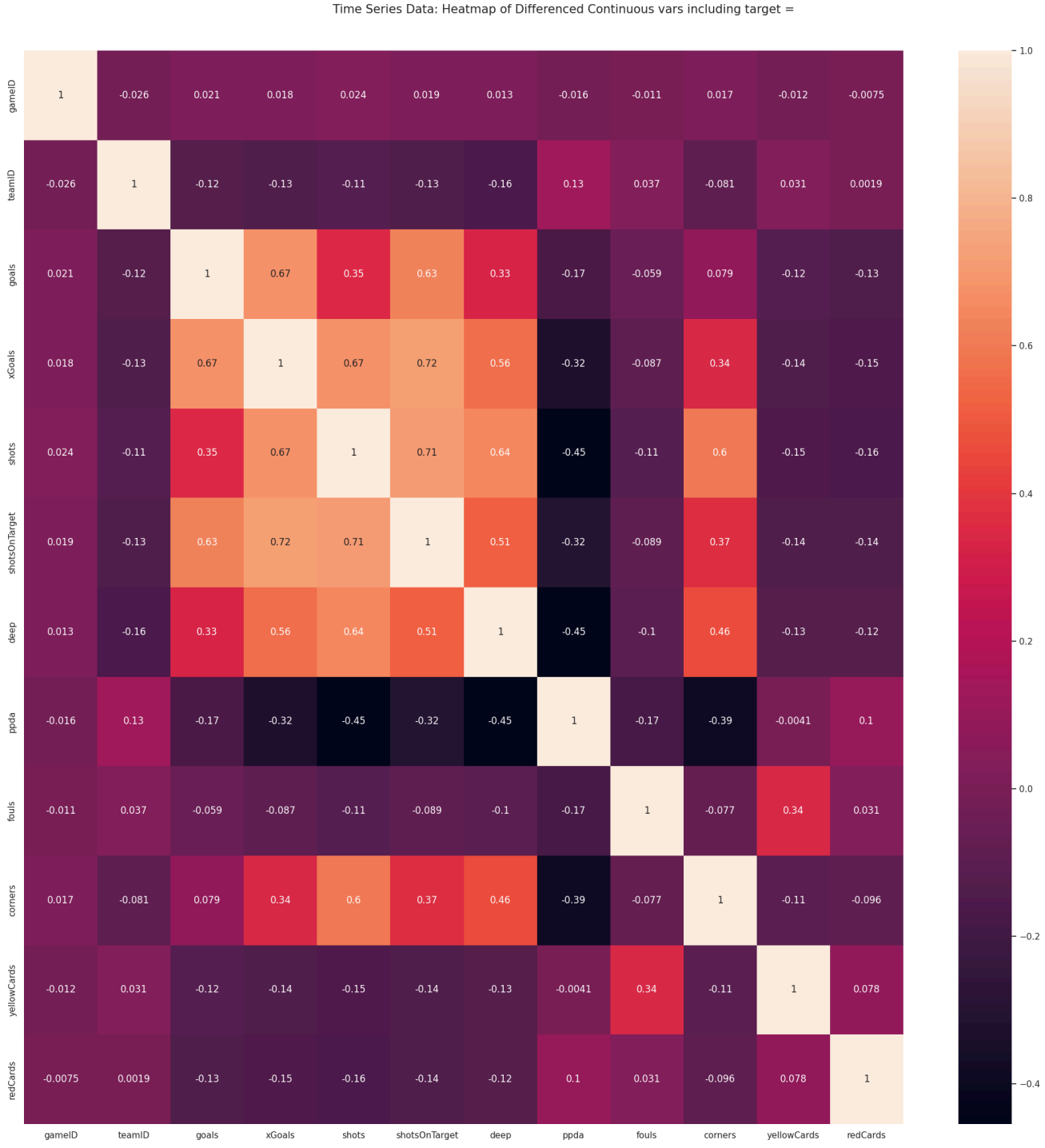

Histograms and Normalized distribitions of all variables
Norm. disti.of location (top 15 categories only)
h
a
Norm. disti.of result (top 15 categories only)
W
L
D
Norm. disti.of gameID (top 15 categories only)
1e−5
81
10339
10341
10342
10343
10344
10345
10346
10347
10348
10349
10350
10351
10352
10353
Norm. disti.of teamID (top 15 categories only)
89
105
95
96
98
99
146
101
102
104
106
150
107
82
143
Norm. disti.of goals (top 15 categories only)
1
0
2
3
4
5
6
7
8
9
10
Norm. disti.of shots (top 15 categories only)
11
10
9
12
13
14
8
15
7
16
6
17
18
5
19
Norm. disti.of shotsOnTarget (top 15 categories only)
3
4
2
5
6
1
7
8
9
0
10
11
12
13
14
Norm. disti.of deep (top 15 categories only)
4
3
5
2
6
7
1
8
9
10
11
0
12
13
14
Norm. disti.of fouls (top 15 categories only)
Norm. disti.of corners (top 15 categories only)


0.02
0.00
12 13 11 14 10 15 9 16 17 8 18 7 19 6 20

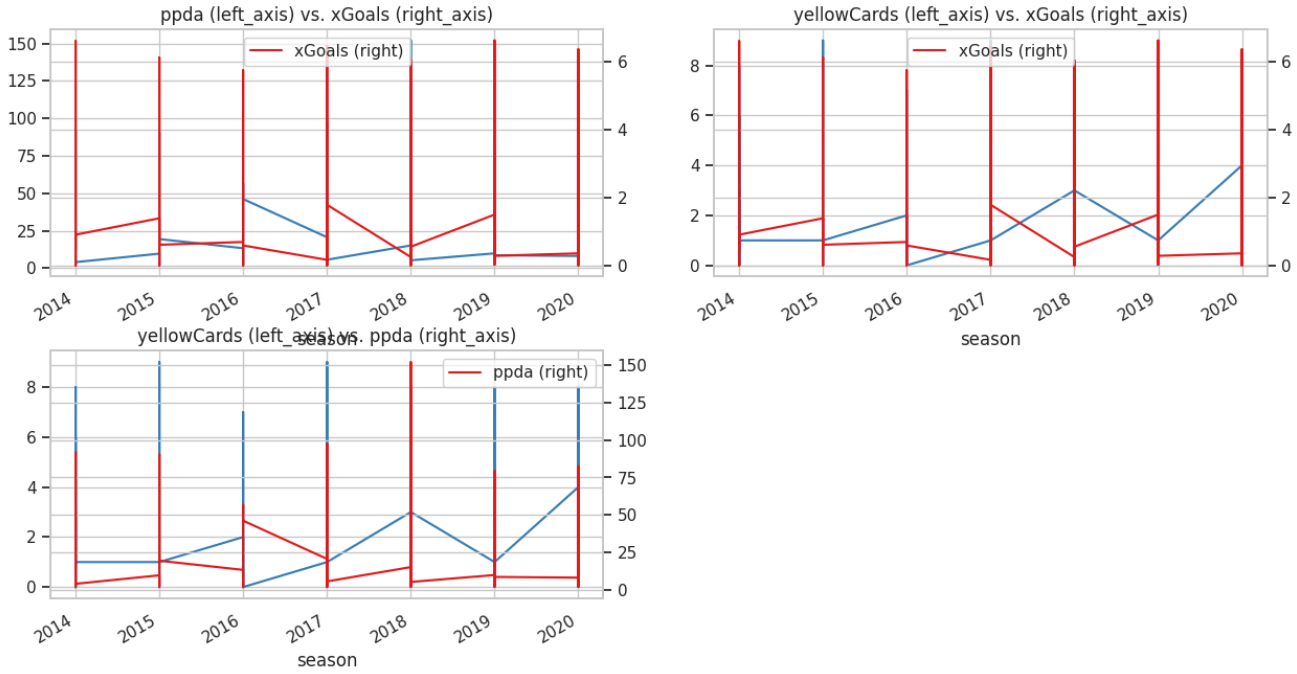

0.04
0.02
0.00
4 3 5 6 2 7 1 8 9 10 0 11 12 13 14


Norm. disti.of redCards (top 15 categories only)
0.8
0.6
0.4
0.2
0.0
0 1 2 3

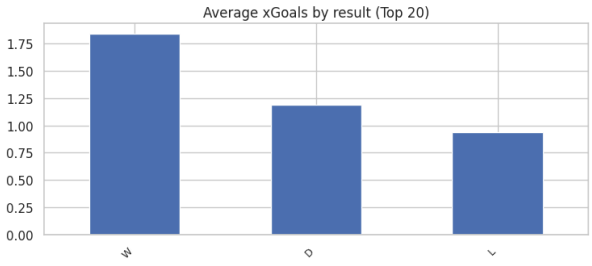

Violin Plot of all Continuous Variables
20
15
10
5
0
xGoals
ppda
yellowCards

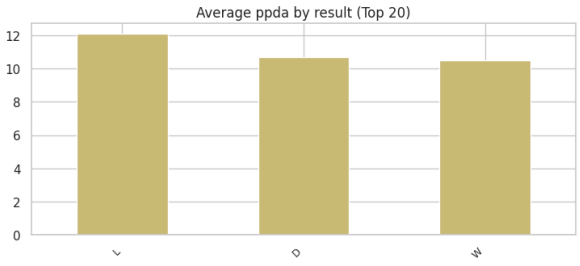

Time Series Data: Heatmap of Differenced Continuous vars including target =
gameID 1 -0.026 0.021 0.018 0.024 0.019 0.013 -0.016 -0.011 0.017 -0.012 -0.0075
teamID -0.026 1 -0.12 -0.13 -0.11 -0.13 -0.16 0.13 0.037 -0.081 0.031 0.0019
goals 0.021 -0.12 1 0.67 0.35 0.63 0.33 -0.17 -0.059 0.079 -0.12 -0.13
xGoals 0.018 -0.13 0.67 1 0.67 0.72 0.56 -0.32 -0.087 0.34 -0.14 -0.15
shots 0.024 -0.11 0.35 0.67 1 0.71 0.64 -0.45 -0.11 0.6 -0.15 -0.16
shotsOnTarget 0.019 -0.13 0.63 0.72 0.71 1 0.51 -0.32 -0.089 0.37 -0.14 -0.14
deep 0.013 -0.16 0.33 0.56 0.64 0.51 1 -0.45 -0.1 0.46 -0.13 -0.12
ppda -0.016 0.13 -0.17 -0.32 -0.45 -0.32 -0.45 1 -0.17 -0.39 -0.0041 0.1
fouls -0.011 0.037 -0.059 -0.087 -0.11 -0.089 -0.1 -0.17 1 -0.077 0.34 0.031
corners 0.017 -0.081 0.079 0.34 0.6 0.37 0.46 -0.39 -0.077 1 -0.11 -0.096
yellowCards -0.012 0.031 -0.12 -0.14 -0.15 -0.14 -0.13 -0.0041 0.34 -0.11 1 0.078
redCards -0.0075 0.0019 -0.13 -0.15 -0.16 -0.14 -0.12 0.1 0.031 -0.096 0.078 1
gameID teamID goals xGoals shots shotsOnTarget deep ppda fouls corners yellowCards redCards
1.0 0.8 0.6 0.4 0.2 0.0 −0.2 −0.4

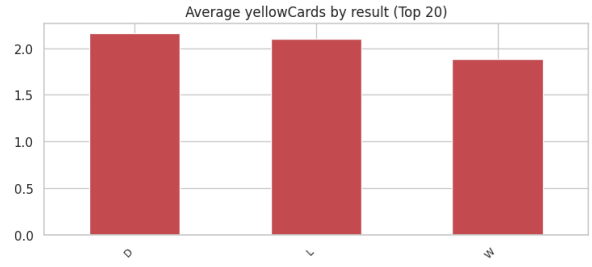
Time Series Plot by season: Pairwise Continuous Variables

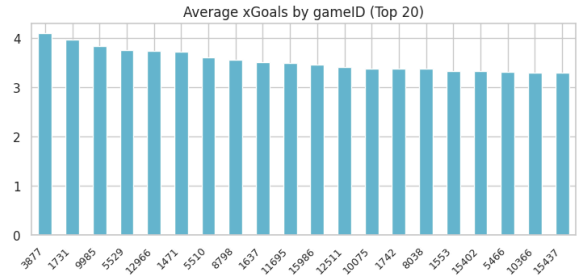


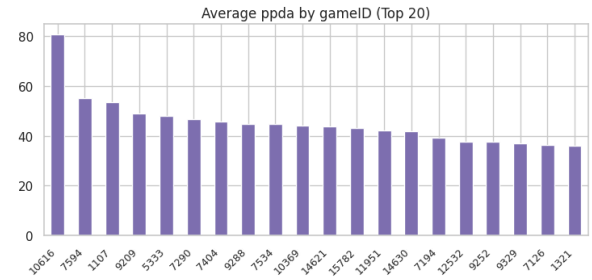
Bar plots for each Continuous by each Categorical variable

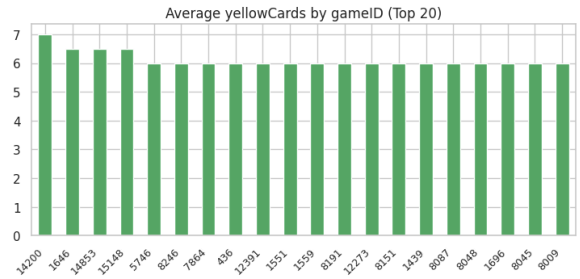


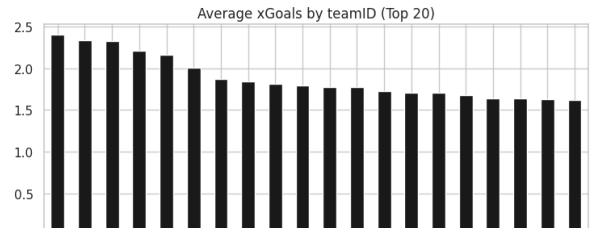


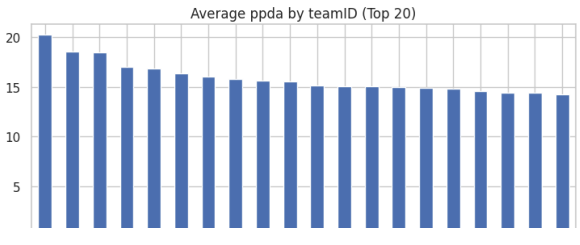


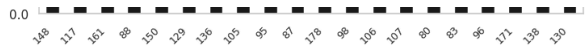


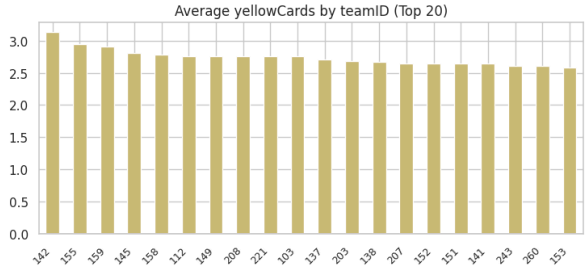


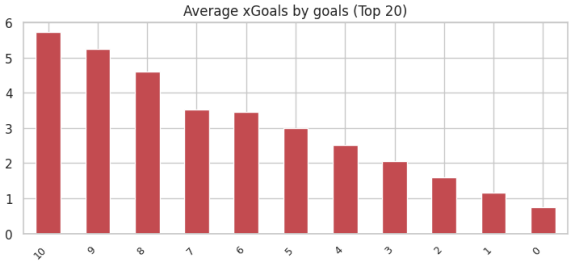


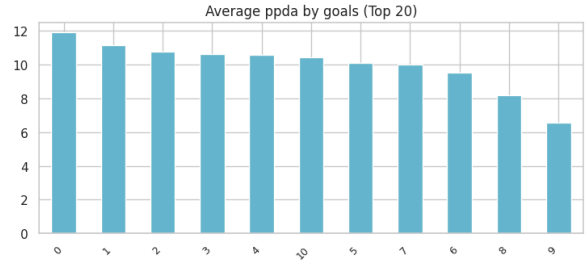


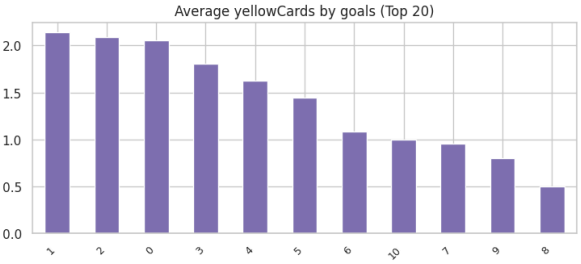


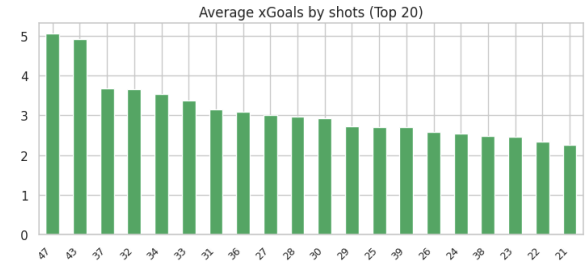


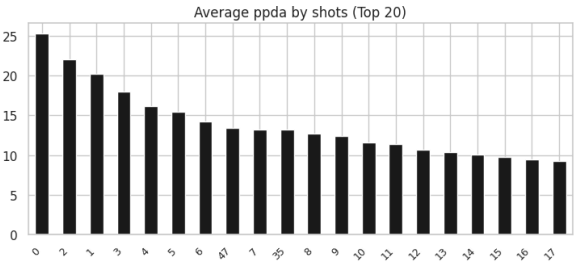

Average yellowCards by teamID (Top 20)

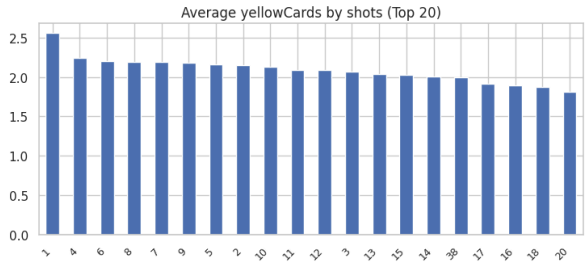

Average xGoals by goals (Top 20)

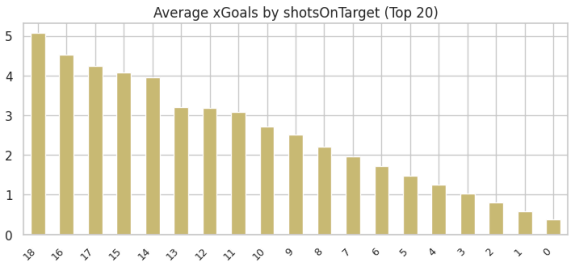

Average ppda by goals (Top 20)

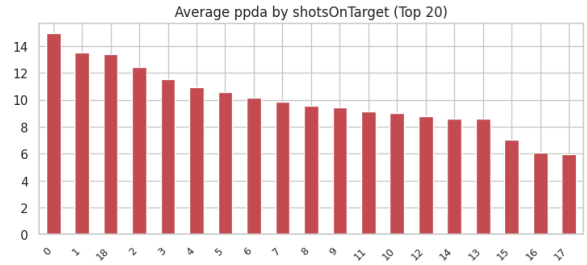

Average yellowCards by goals (Top 20)

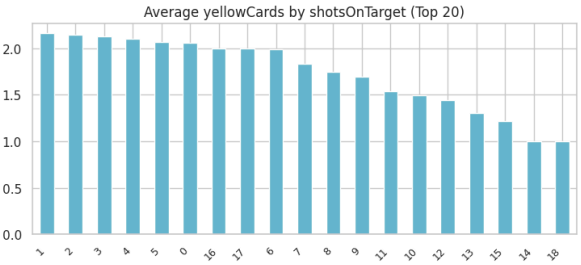

Average xGoals by shots (Top 20)

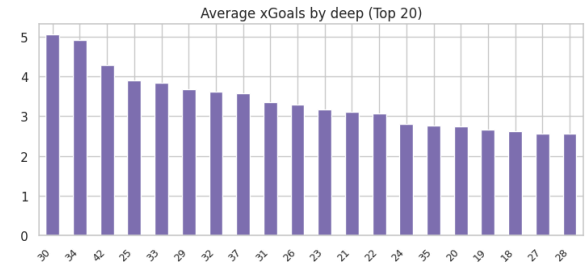

Average ppda by shots (Top 20)

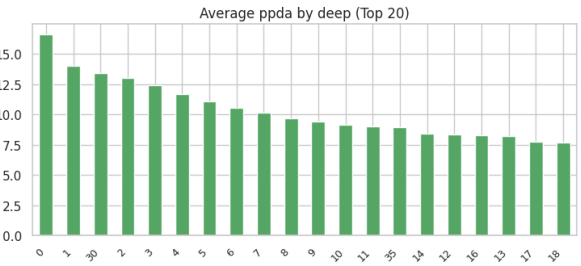

Average yellowCards by shots (Top 20)


Average xGoals by shotsOnTarget (Top 20)


Average ppda by shotsOnTarget (Top 20)

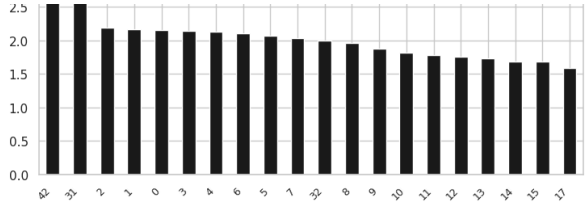

Average yellowCards by shotsOnTarget (Top 20)

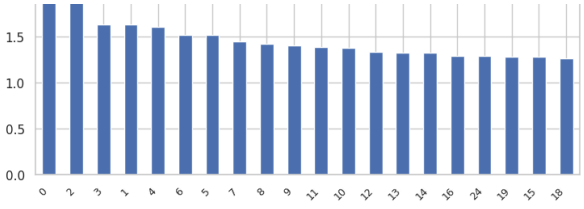

Average xGoals by deep (Top 20)

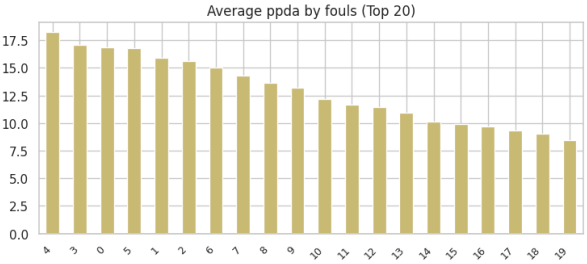

Average ppda by deep (Top 20)

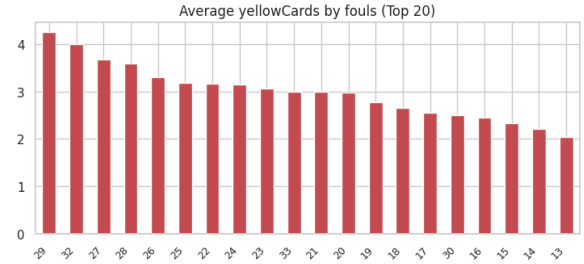

Average yellowCards by deep (Top 20)

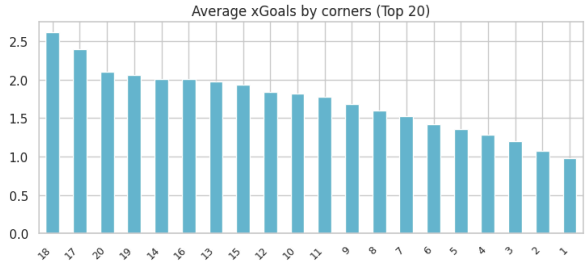

Average xGoals by fouls (Top 20)

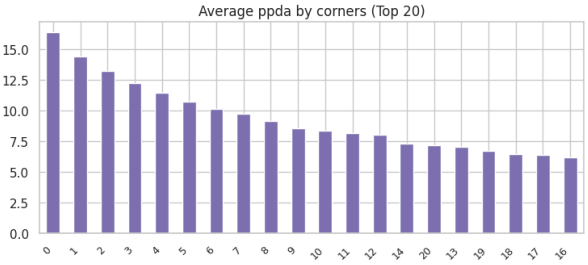


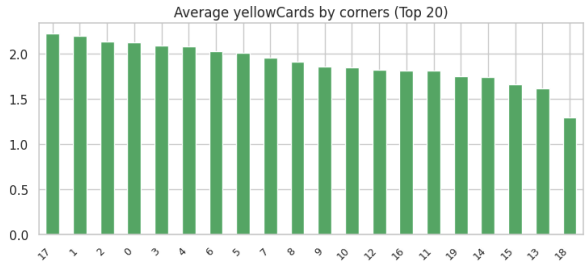


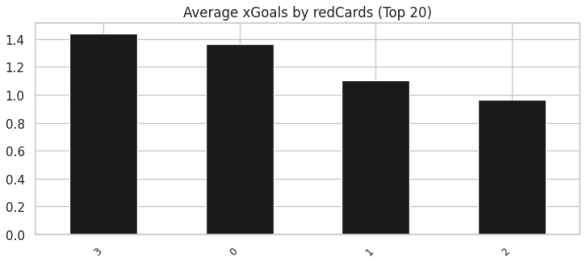

Average ppda by fouls (Top 20)

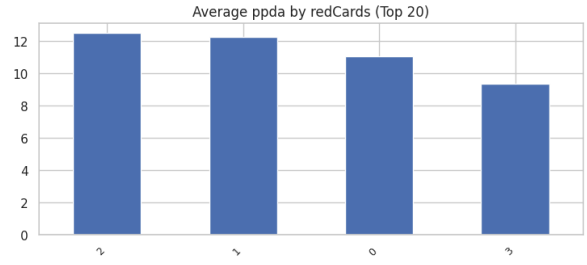

Average yellowCards by fouls (Top 20)

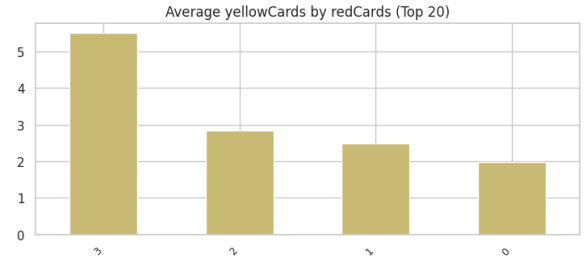

Average xGoals by corners (Top 20)

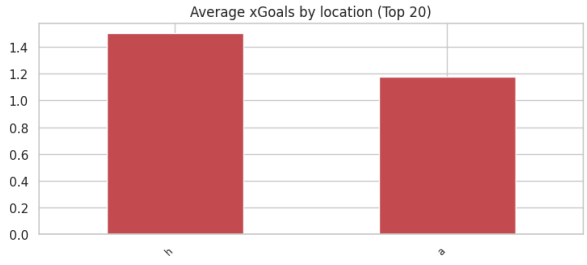

Average ppda by corners (Top 20)

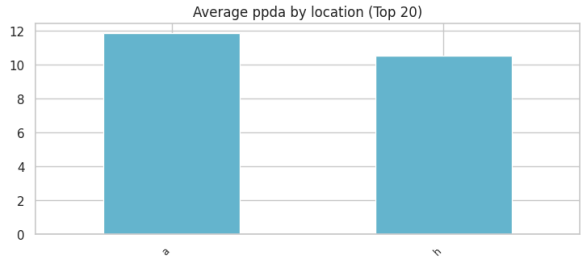

Average yellowCards by corners (Top 20)

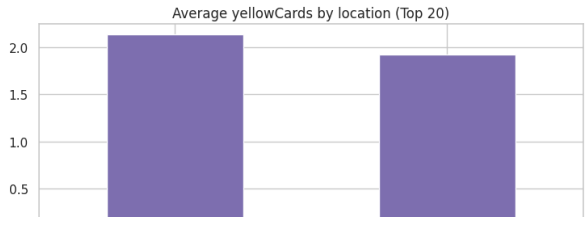

Average xGoals by redCards (Top 20)


Average ppda by redCards (Top 20)


Average yellowCards by redCards (Top 20)


Average xGoals by location (Top 20)


Average ppda by location (Top 20)


Average yellowCards by location (Top 20)

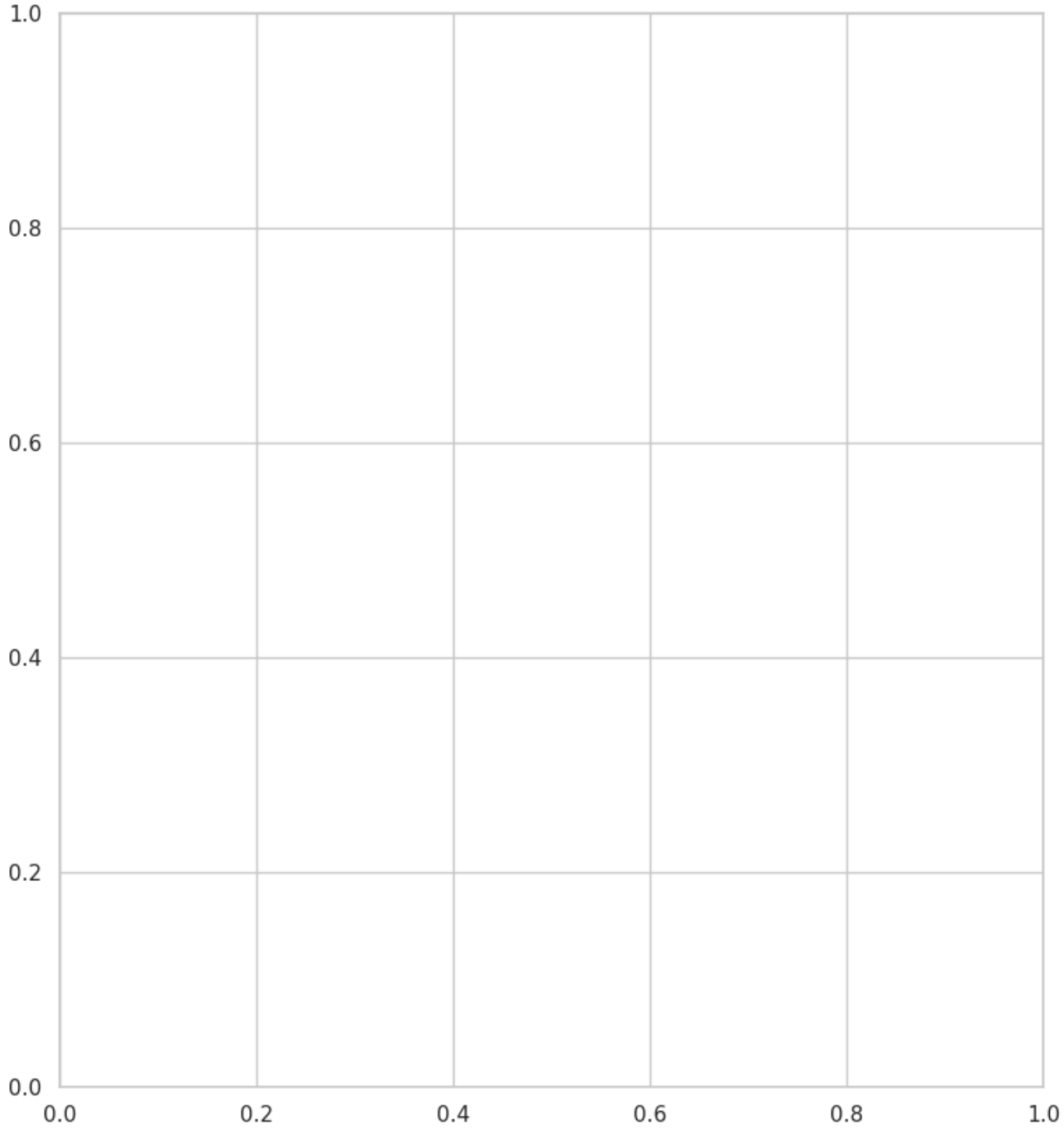

0.0

```
[nltk_data] Downloading collection 'popular'
[nltk_data]    |
[nltk_data]    | Downloading package cmudict to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package cmudict is already up-to-date!
[nltk_data]    | Downloading package gazetteers to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package gazetteers is already up-to-date!
[nltk_data]    | Downloading package genesis to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package genesis is already up-to-date!
[nltk_data]    | Downloading package gutenberg to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package gutenberg is already up-to-date!
[nltk_data]    | Downloading package inaugural to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package inaugural is already up-to-date!
[nltk_data]    | Downloading package movie_reviews to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package movie_reviews is already up-to-date!
[nltk_data]    | Downloading package names to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package names is already up-to-date!
[nltk_data]    | Downloading package shakespeare to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package shakespeare is already up-to-date!
[nltk_data]    | Downloading package stopwords to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package stopwords is already up-to-date!
[nltk_data]    | Downloading package treebank to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package treebank is already up-to-date!
[nltk_data]    | Downloading package twitter_samples to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package twitter_samples is already up-to-date!
[nltk_data]    | Downloading package omw to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package omw is already up-to-date!
[nltk_data]    | Downloading package omw-1.4 to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package omw-1.4 is already up-to-date!
[nltk_data]    | Downloading package wordnet to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package wordnet is already up-to-date!
[nltk_data]    | Downloading package wordnet2021 to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package wordnet2021 is already up-to-date!
[nltk_data]    | Downloading package wordnet31 to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package wordnet31 is already up-to-date!
[nltk_data]    | Downloading package wordnet_ic to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package wordnet_ic is already up-to-date!
[nltk_data]    | Downloading package words to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package words is already up-to-date!
[nltk_data]    | Downloading package maxent_ne_chunker to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package maxent_ne_chunker is already up-to-date!
[nltk_data]    | Downloading package punkt to
```

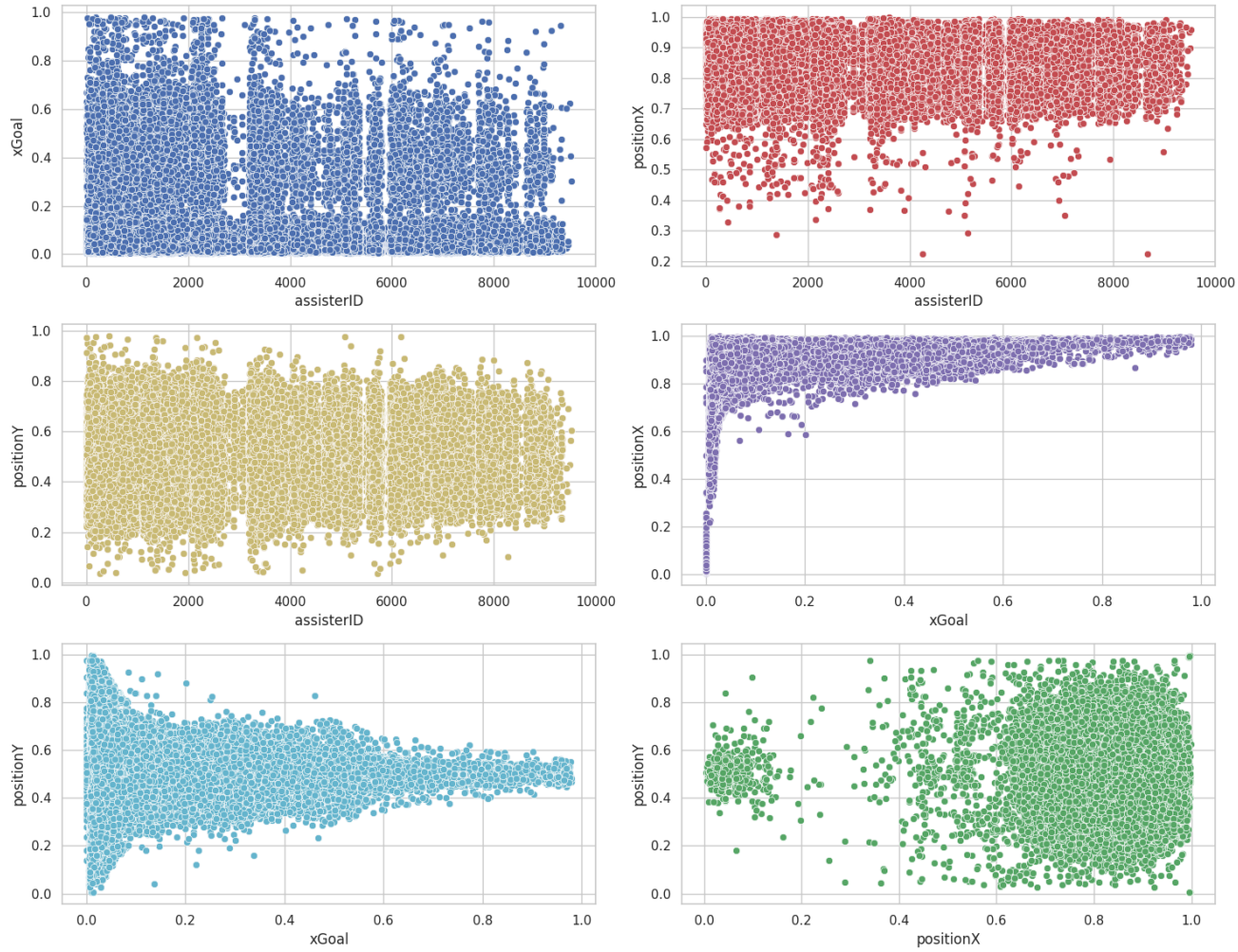
```
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package punkt is already up-to-date!
[nltk_data]    | Downloading package snowball_data to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package snowball_data is already up-to-date!
[nltk_data]    | Downloading package averaged_perceptron_tagger to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package averaged_perceptron_tagger is already up-
[nltk_data]    |       to-date!
[nltk_data]    |
[nltk_data]  Done downloading collection popular
Could not draw wordcloud plot for date
All Plots done
Time to run AutoViz = 20 seconds

 ###################### AUTO VISUALIZATION Completed ######################
=== Completed AutoViz on df_teamstats ===

=== Running AutoViz on df_shots ===
    max_rows_analyzed is smaller than dataset shape 324543...
        randomly sampled 150000 rows from read CSV file
Shape of your Data Set loaded: (150000, 11)
################################################################################
#####
######################## C L A S S I F Y I N G  V A R I A B L E S  ###############
#####
################################################################################
#####
Classifying variables in data set...
    Number of Numeric Columns =  4
    Number of Integer-Categorical Columns =  3
    Number of String-Categorical Columns =  4
    Number of Factor-Categorical Columns =  0
    Number of String-Boolean Columns =  0
    Number of Numeric-Boolean Columns =  0
    Number of Discrete String Columns =  0
    Number of NLP String Columns =  0
    Number of Date Time Columns =  0
    Number of ID Columns =  0
    Number of Columns to Delete =  0
    11 Predictors classified...
        No variables removed since no ID or low-information variables found in dat
a set
Since Number of Rows in data 150000 exceeds maximum, randomly sampling 150000 rows
for EDA...
To fix these data quality issues in the dataset, import FixDQ from autoviz...
    All variables classified into correct types.
```

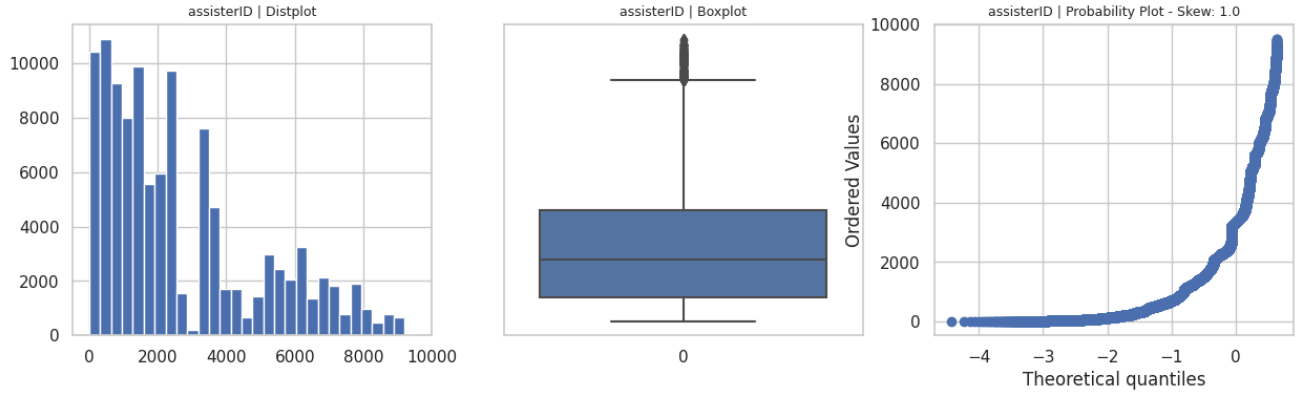
	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ Issue
gameID	int64	0.000000	8	81.000000	16135.000000	No issue
shooterID	int64	0.000000	3	3.000000	9566.000000	Column has 2332 outliers greater than upper bound (8225.50) or lower than lower bound(-3578.50). Cap them or remove them.
assisterID	float64	26.037333	NA	1.000000	9526.000000	39056 missing values. Impute them with mean, median, mode, or a constant value such as 123., Column has 2133 outliers greater than upper bound (8127.00) or lower than lower bound(-3553.00). Cap them or remove them.
minute	int64	0.000000	0	0.000000	103.000000	No issue
situation	object	0.000000	0			No issue
lastAction	object	0.000000	0			24 rare categories: Too many to list. Group them into a single category or drop the categories.
shotType	object	0.000000	0			1 rare categories: ['OtherBodyPart']. Group them into a single category or drop the categories.
shotResult	object	0.000000	0			1 rare categories: ['OwnGoal']. Group them into a single category or drop the categories.
xGoal	float64	0.000000	NA	0.000000	0.979344	Column has 20996 outliers greater than upper bound (0.20) or lower than lower bound(-0.08). Cap them or remove them.
positionX	float64	0.000000	NA	0.004000	0.999000	Column has 867 outliers greater than upper bound (1.10) or lower than lower bound(0.59). Cap them or remove them.
positionY	float64	0.000000	NA	0.005000	0.997000	Column has 322 outliers greater than upper bound (0.87) or lower than lower bound(0.14). Cap them or remove them.

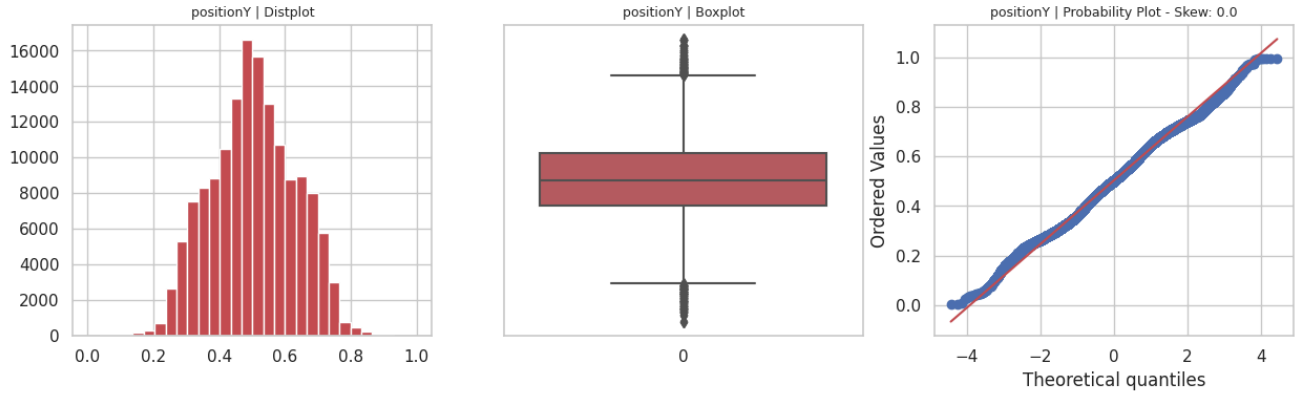
```
Number of All Scatter Plots = 10
```

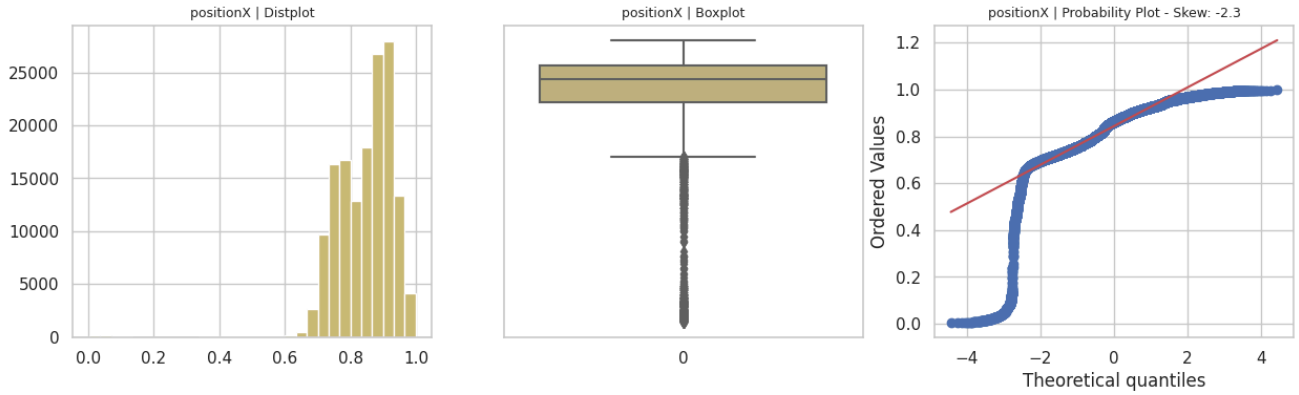

1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0

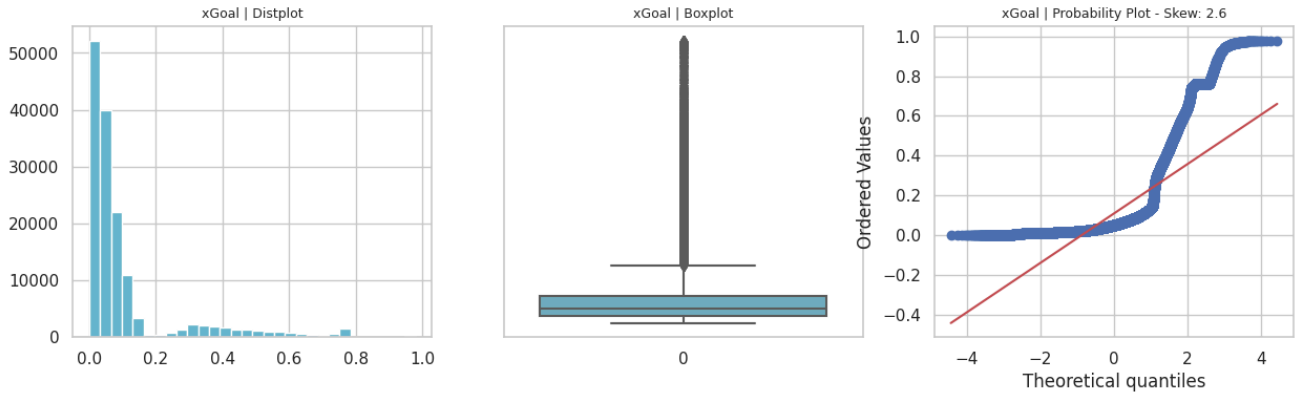
Pair-wise Scatter Plot of all Continuous Variables

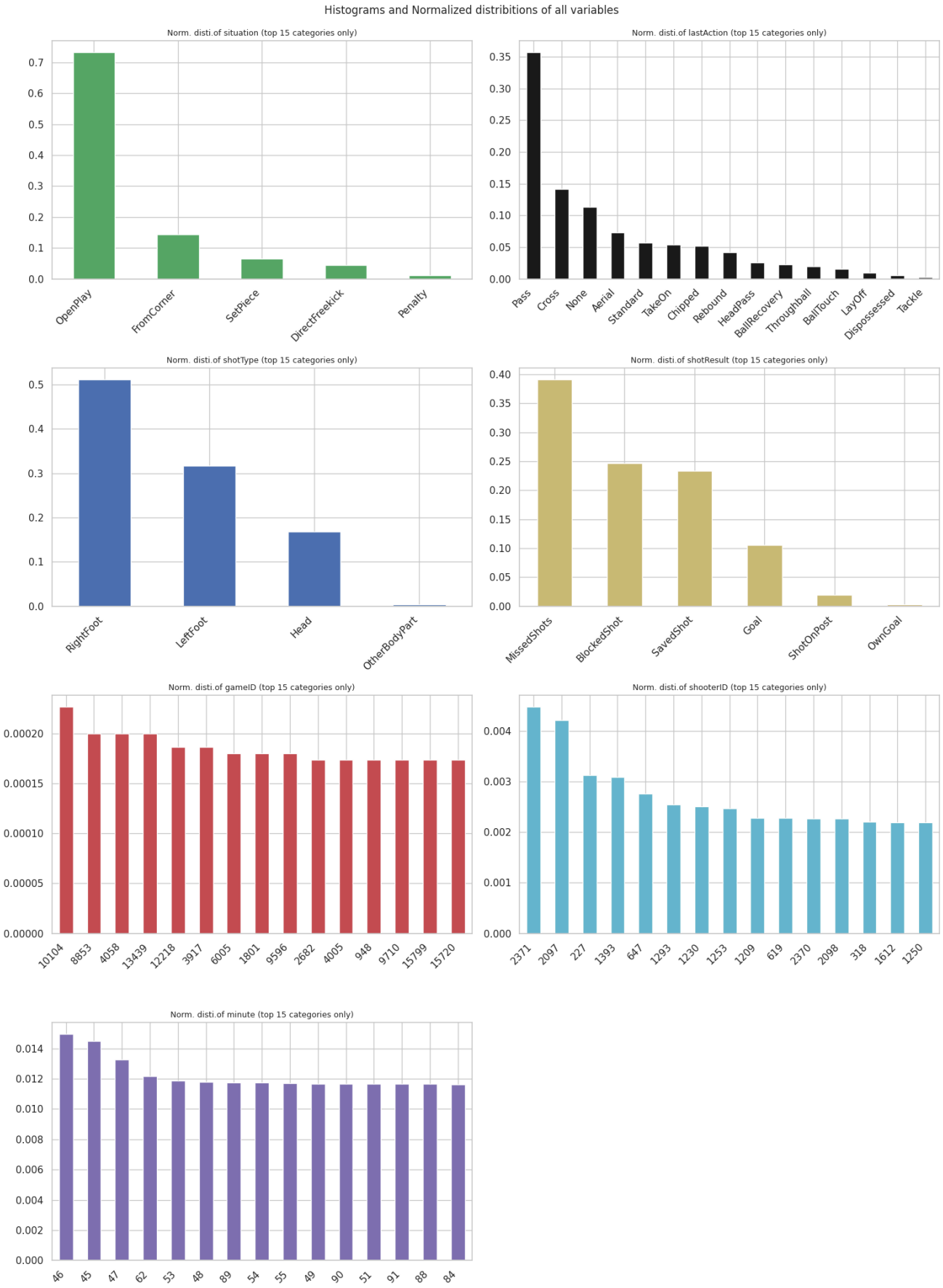



assisterID | Distplot
assisterID | Boxplot
assisterID | Probability Plot - Skew: 1.0
Ordered Values
Theoretical quantiles

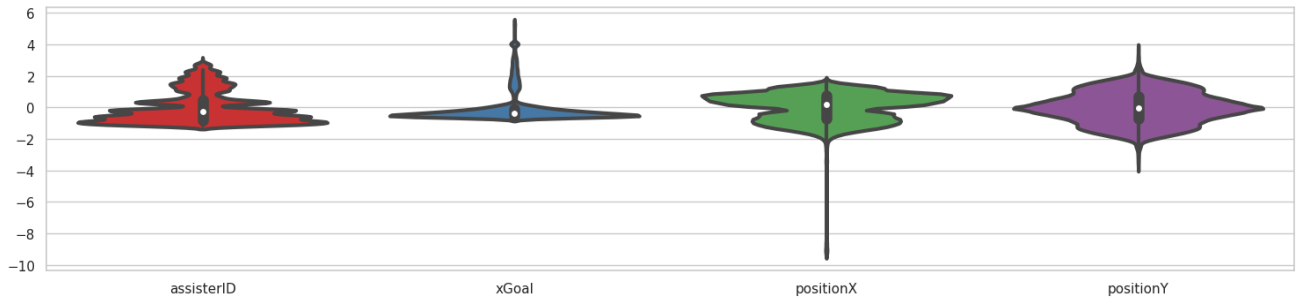

positionY | Distplot
positionY | Boxplot
positionY | Probability Plot - Skew: 0.0
Ordered Values
Theoretical quantiles


positionX | Distplot
positionX | Boxplot
positionX | Probability Plot - Skew: -2.3
Ordered Values
Theoretical quantiles

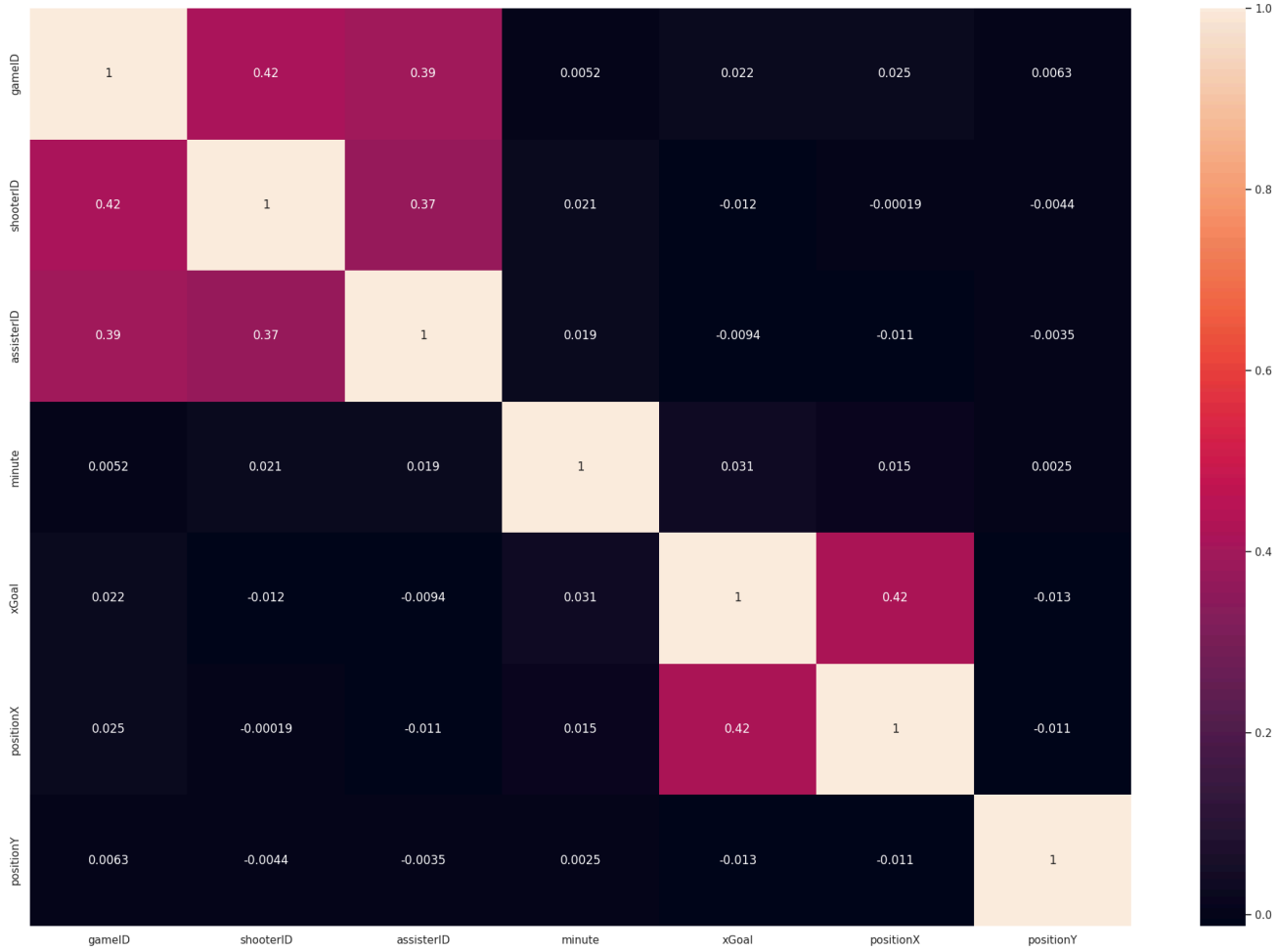

xGoal | Distplot
xGoal | Boxplot
xGoal | Probability Plot - Skew: 2.6
Ordered Values
Theoretical quantiles


Histograms and Normalized distribitions of all variables
Norm. disti.of situation (top 15 categories only)
OpenPlay
FromCorner
SetPiece
DirectFreekick
Penalty
Norm. disti.of lastAction (top 15 categories only)
Pass
Cross
None
Aerial
Standard
TakeOn
Chipped
Rebound
HeadPass
BallRecovery
Throughball
BallTouch
LayOff
Dispossessed
Tackle
Norm. disti.of shotType (top 15 categories only)
RightFoot
LeftFoot
Head
OtherBodyPart
Norm. disti.of shotResult (top 15 categories only)
MissedShots
BlockedShot
SavedShot
Goal
ShotOnPost
OwnGoal
Norm. disti.of gameID (top 15 categories only)
10104
8853
4058
13439
12218
3917
6005
1801
9596
2682
4005
948
9710
15799
15720
Norm. disti.of shooterID (top 15 categories only)
2371
2097
227
1393
647
1293
1230
1253
1209
619
2370
2098
318
1612
1250
Norm. disti.of minute (top 15 categories only)
46
45
47
62
53
48
89
54
55
49
90
51
91
88
84

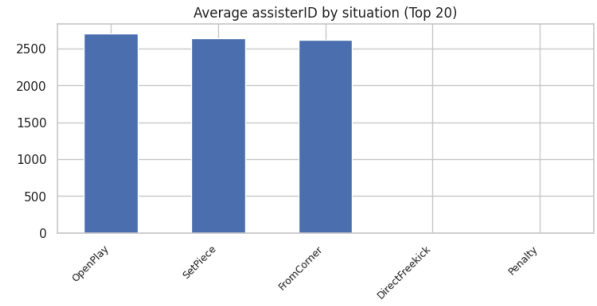
Violin Plot of all Continuous Variables

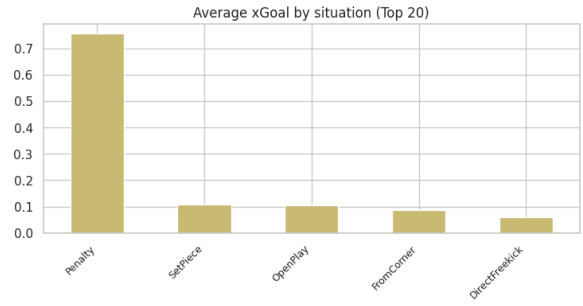


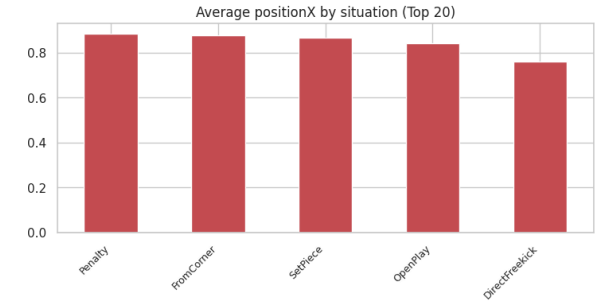
Heatmap of all Numeric Variables including target:

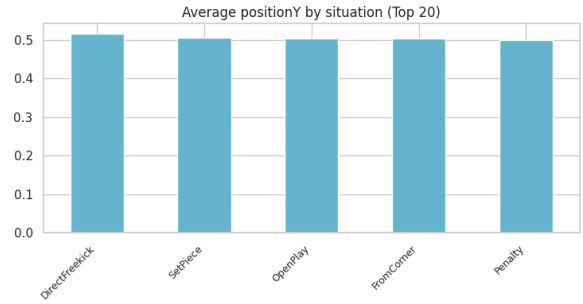


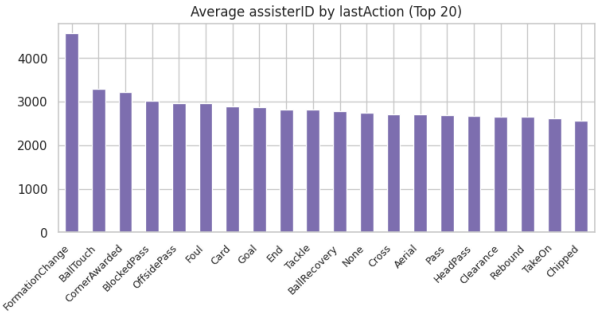
Bar plots for each Continuous by each Categorical variable

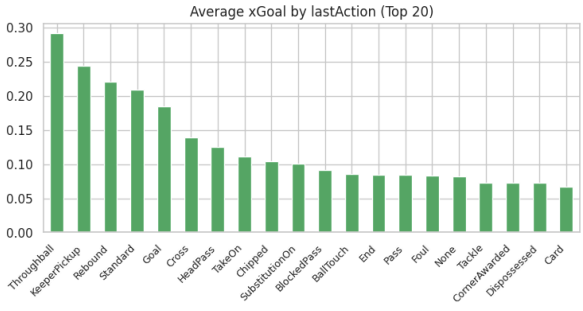


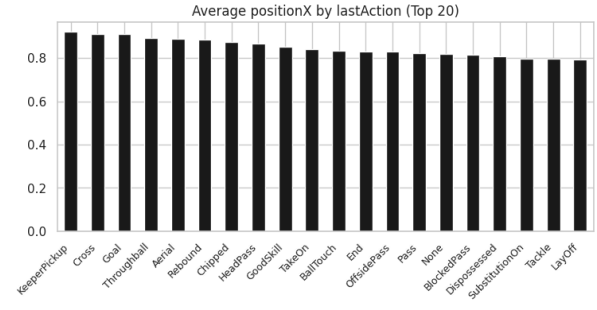


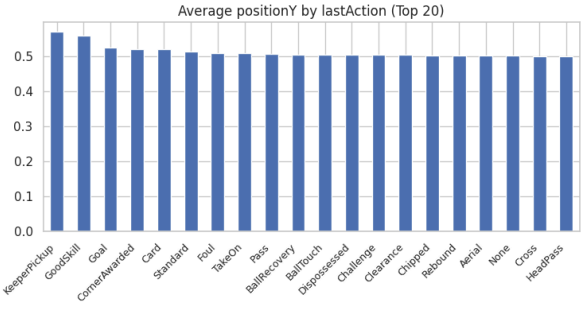


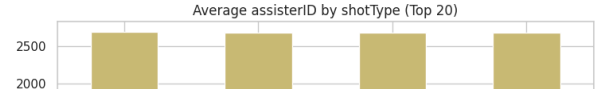


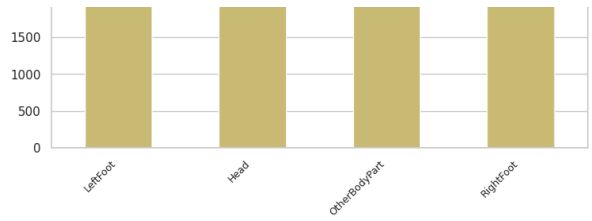


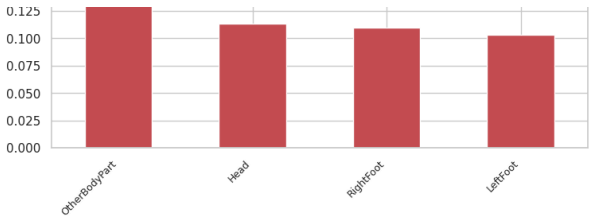


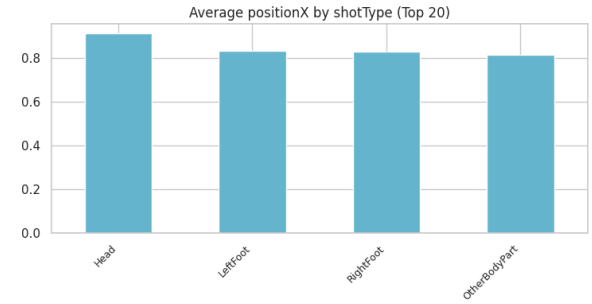


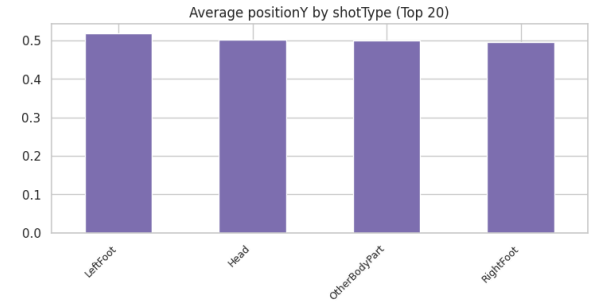


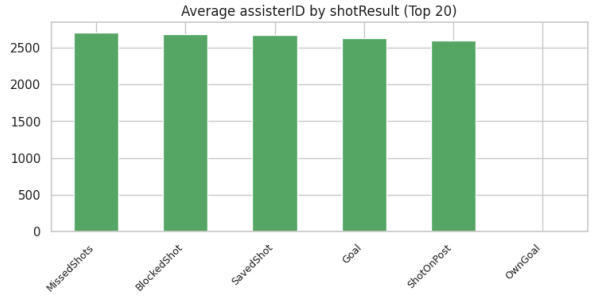


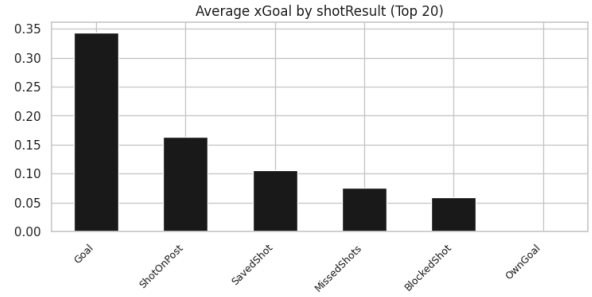

1500
1000
500
0
LeftFoot
Head
OtherBodyPart
RightFoot

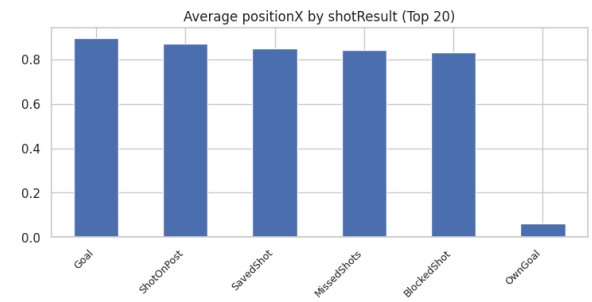

0.125
0.100
0.075
0.050
0.025
0.000
OtherBodyPart
Head
RightFoot
LeftFoot

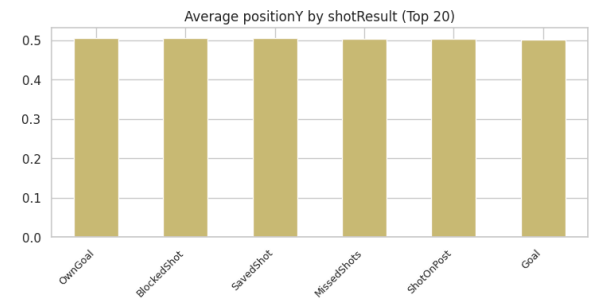

Average positionX by shotType (Top 20)
0.8
0.6
0.4
0.2
0.0
Head
LeftFoot
RightFoot
OtherBodyPart

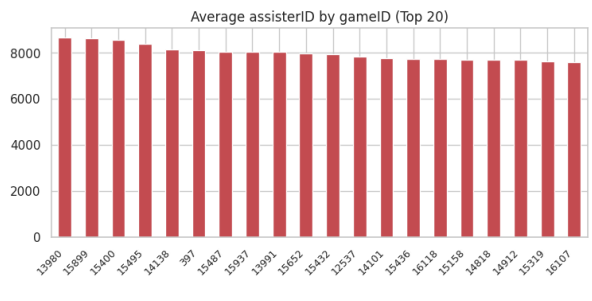

Average positionY by shotType (Top 20)
0.5
0.4
0.3
0.2
0.1
0.0
LeftFoot
Head
OtherBodyPart
RightFoot

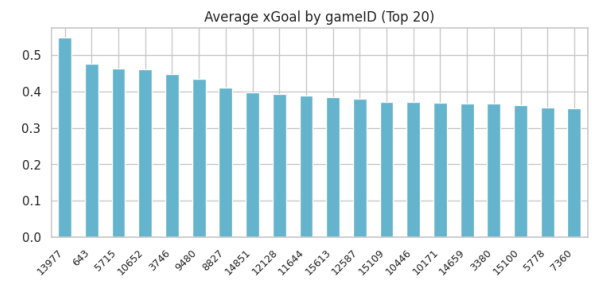

Average assisterID by shotResult (Top 20)
2500
2000
1500
1000
500
0
MissedShots
BlockedShot
SavedShot
Goal
ShotOnPost
OwnGoal

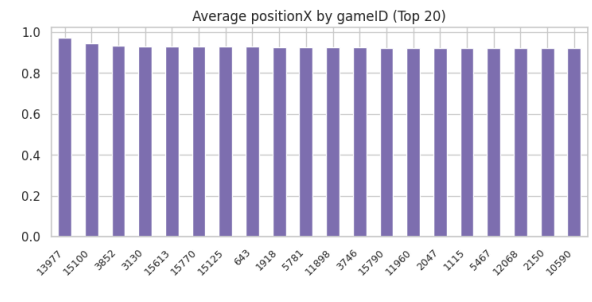

Average xGoal by shotResult (Top 20)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Goal
ShotOnPost
SavedShot
MissedShots
BlockedShot
OwnGoal

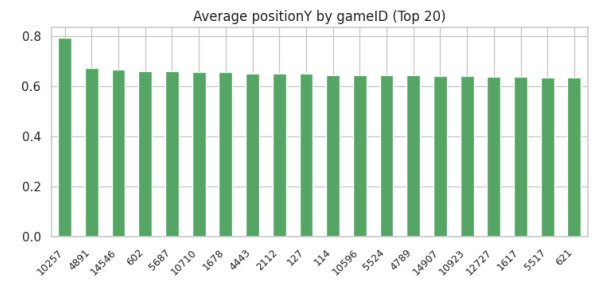

Average positionX by shotResult (Top 20)
0.8
0.6
0.4
0.2
0.0
Goal
ShotOnPost
SavedShot
MissedShots
BlockedShot
OwnGoal

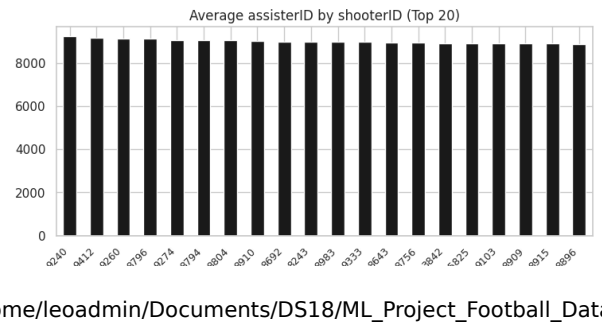

Average positionY by shotResult (Top 20)
0.5
0.4
0.3
0.2
0.1
0.0
OwnGoal
BlockedShot
SavedShot
MissedShots
ShotOnPost
Goal

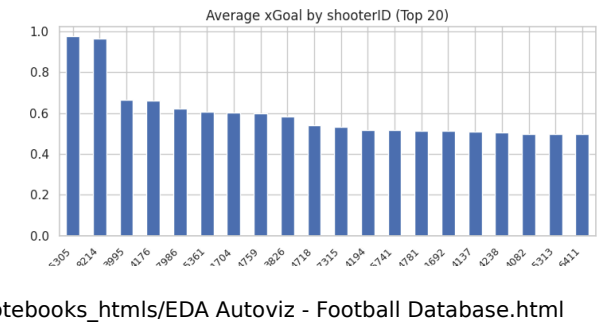

Average assisterID by gameID (Top 20)
8000
6000
4000
2000
0
13960
15899
15400
15495
14138
397
15487
15937
13991
15652
15432
12537
14101
15436
16118
15158
14818
14912
15319
16107

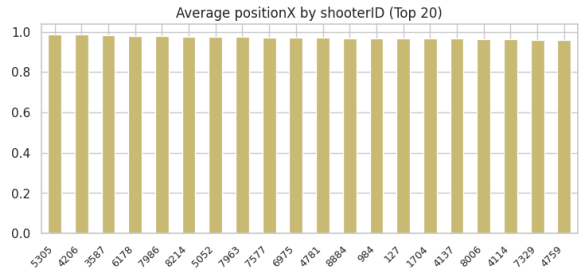

Average xGoal by gameID (Top 20)
0.5
0.4
0.3
0.2
0.1
0.0
13977
643
5715
10652
3746
9480
8827
14851
12128
11644
15613
12587
15109
10446
10171
14659
3380
15100
5778
7360

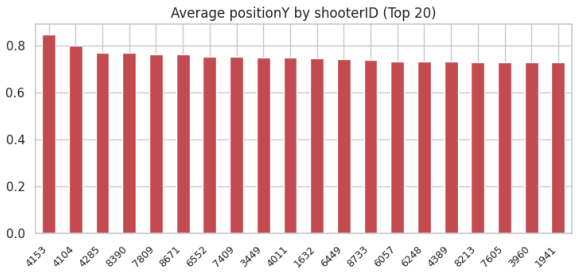

Average positionX by gameID (Top 20)
1.0
0.8
0.6
0.4
0.2
0.0
13977
15100
3852
3130
15613
15770
15125
643
1918
5781
11898
3746
15790
11960
2047
1115
5467
12068
2150
10590

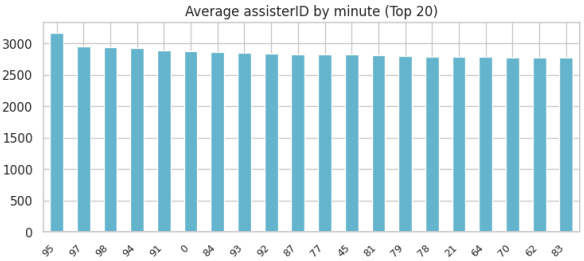

Average positionY by gameID (Top 20)
0.8
0.6
0.4
0.2
0.0
10257
4891
14546
602
5687
10710
1678
4443
2112
127
114
10596
5524
4789
14907
10923
12727
1617
5517
621

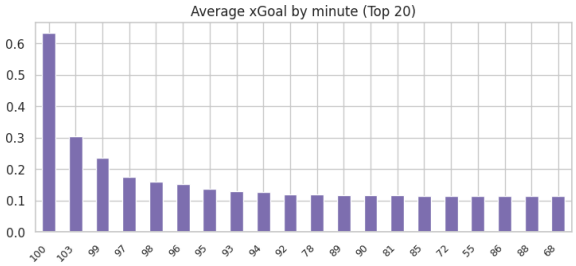

Average assisterID by shooterID (Top 20)
8000
6000
4000
2000
0

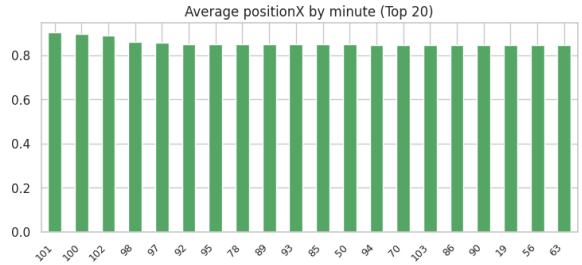

Average xGoal by shooterID (Top 20)
1.0
0.8
0.6
0.4
0.2
0.0

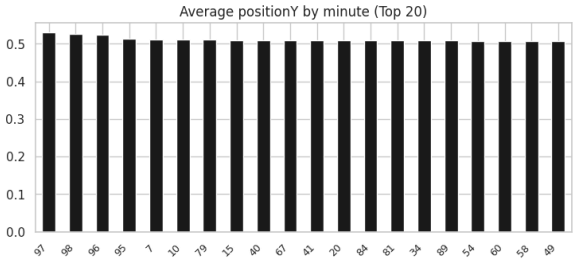

Average positionX by shooterID (Top 20)
1.0
0.8
0.6
0.4
0.2
0.0
5305 4206 3587 6178 7986 8214 5052 7963 7577 6975 4781 8884 984 127 1704 4137 8006 4114 7329 4759


Average positionY by shooterID (Top 20)
0.8
0.6
0.4
0.2
0.0
4153 4104 4285 8390 7809 8671 6552 7409 3449 4011 1632 6449 8733 6057 6248 4389 8213 7605 3960 1941


Average assisterID by minute (Top 20)
3000
2500
2000
1500
1000
500
0
95 97 98 94 91 0 84 93 92 87 77 45 81 79 78 21 64 70 62 83


Average xGoal by minute (Top 20)
0.6
0.5
0.4
0.3
0.2
0.1
0.0
100 103 99 97 98 96 95 93 94 92 78 89 90 81 85 72 55 86 88 68


Average positionX by minute (Top 20)
0.8
0.6
0.4
0.2
0.0
101 100 102 98 97 92 95 78 89 93 85 50 94 70 103 86 90 19 56 63

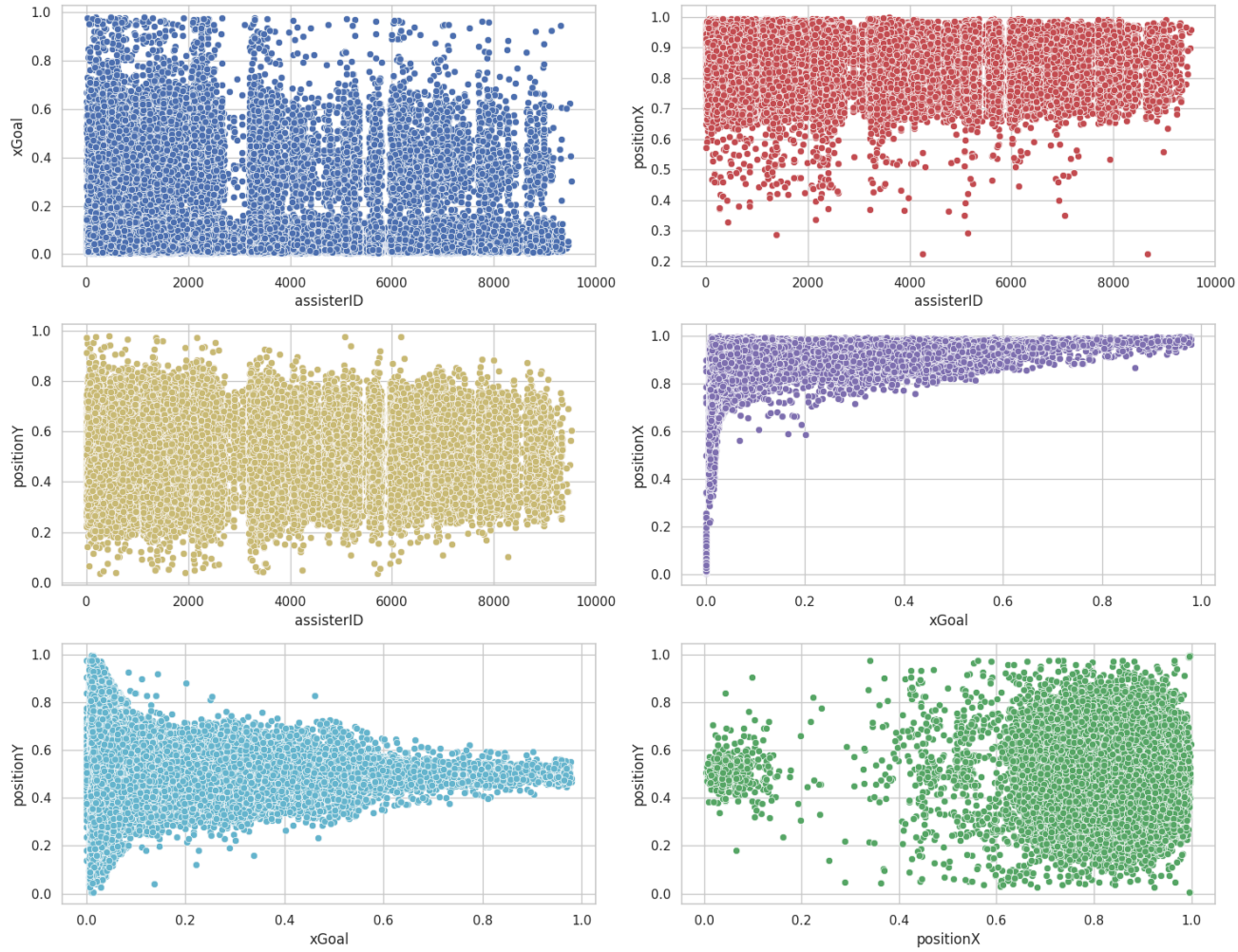

Average positionY by minute (Top 20)
0.5
0.4
0.3
0.2
0.1
0.0
97 98 96 95 7 10 79 15 40 67 41 20 84 81 34 89 54 60 58 49

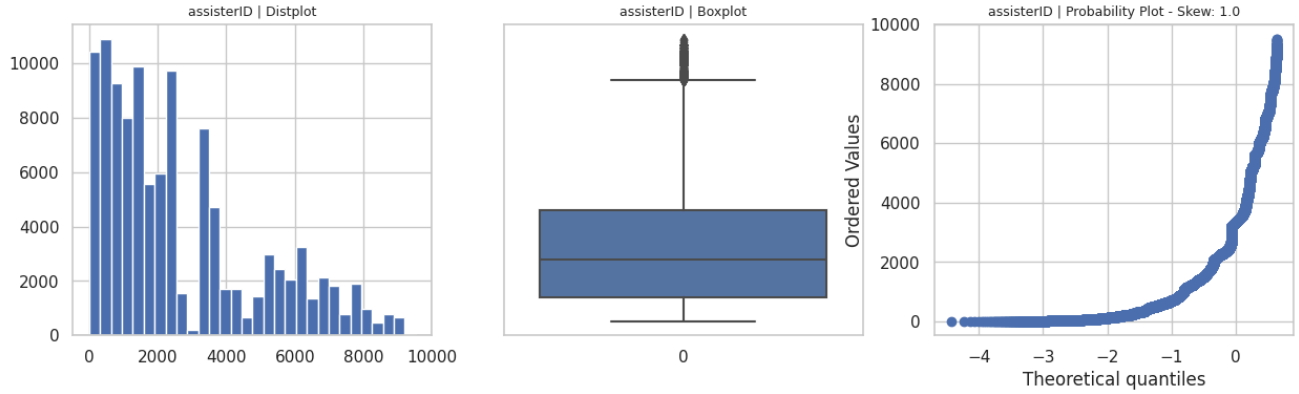
```
All Plots done
Time to run AutoViz = 16 seconds

 ###################### AUTO VISUALIZATION Completed ######################
=== Completed AutoViz on df_shots ===

=== Running AutoViz on player_shots ===
    max_rows_analyzed is smaller than dataset shape 324543...
        randomly sampled 150000 rows from read CSV file
Shape of your Data Set loaded: (150000, 11)
############################################################################################
#####
######################## C L A S S I F Y I N G  V A R I A B L E S  ###############
#####
############################################################################################
#####
Classifying variables in data set...
    Number of Numeric Columns =  4
    Number of Integer-Categorical Columns =  3
    Number of String-Categorical Columns =  4
    Number of Factor-Categorical Columns =  0
    Number of String-Boolean Columns =  0
    Number of Numeric-Boolean Columns =  0
    Number of Discrete String Columns =  0
    Number of NLP String Columns =  0
    Number of Date Time Columns =  0
    Number of ID Columns =  0
    Number of Columns to Delete =  0
    11 Predictors classified...
        No variables removed since no ID or low-information variables found in dat
a set
Since Number of Rows in data 150000 exceeds maximum, randomly sampling 150000 rows
for EDA...
To fix these data quality issues in the dataset, import FixDQ from autoviz...
    All variables classified into correct types.
```

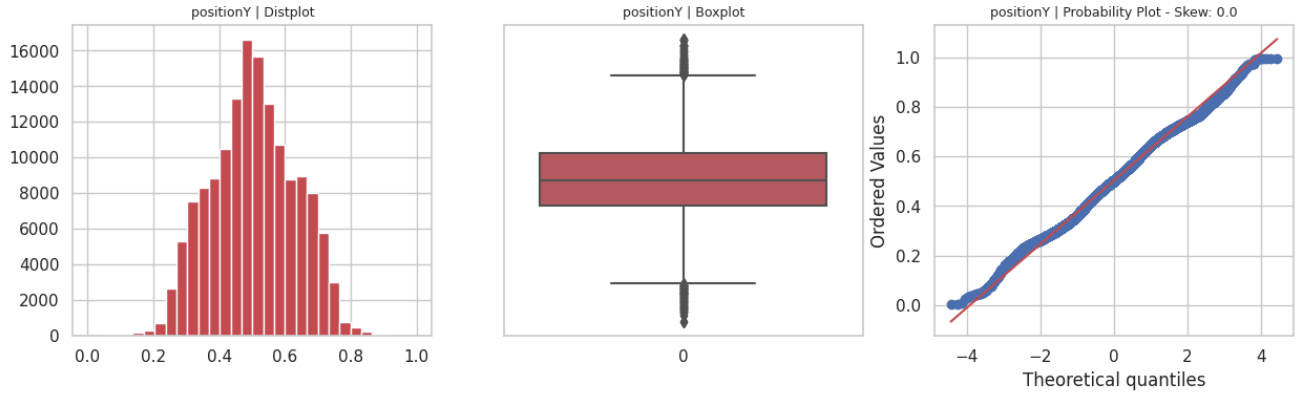
	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ Issue
gameID	int64	0.000000	8	81.000000	16135.000000	No issue
playerID	int64	0.000000	3	3.000000	9566.000000	Column has 2332 outliers greater than upper bound (8225.50) or lower than lower bound(-3578.50). Cap them or remove them.
assisterID	float64	26.037333	NA	1.000000	9526.000000	39056 missing values. Impute them with mean, median, mode, or a constant value such as 123., Column has 2133 outliers greater than upper bound (8127.00) or lower than lower bound(-3553.00). Cap them or remove them.
minute	int64	0.000000	0	0.000000	103.000000	No issue
situation	object	0.000000	0			No issue
lastAction	object	0.000000	0			24 rare categories: Too many to list. Group them into a single category or drop the categories.
shotType	object	0.000000	0			1 rare categories: ['OtherBodyPart']. Group them into a single category or drop the categories.
shotResult	object	0.000000	0			1 rare categories: ['OwnGoal']. Group them into a single category or drop the categories.
xGoal	float64	0.000000	NA	0.000000	0.979344	Column has 20996 outliers greater than upper bound (0.20) or lower than lower bound(-0.08). Cap them or remove them.
positionX	float64	0.000000	NA	0.004000	0.999000	Column has 867 outliers greater than upper bound (1.10) or lower than lower bound(0.59). Cap them or remove them.
positionY	float64	0.000000	NA	0.005000	0.997000	Column has 322 outliers greater than upper bound (0.87) or lower than lower bound(0.14). Cap them or remove them.

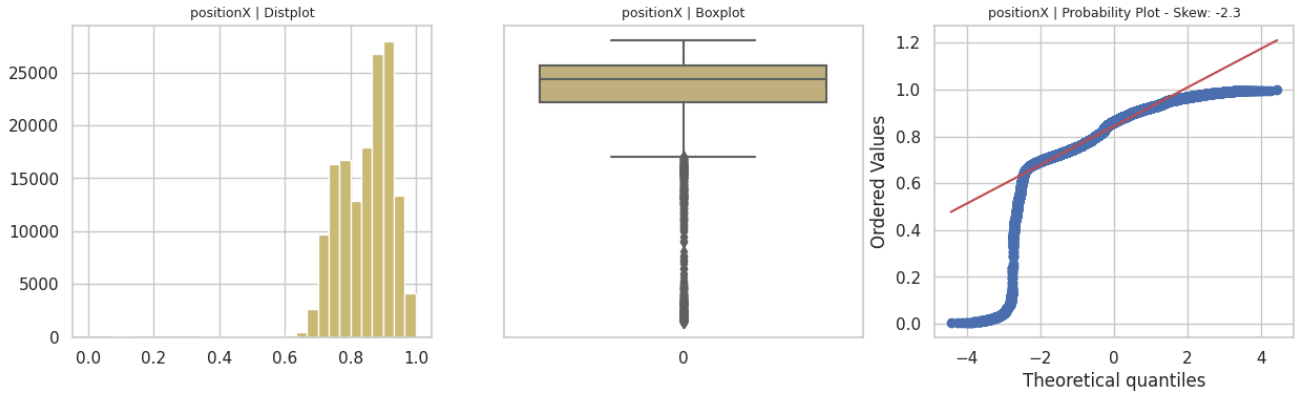
```
Number of All Scatter Plots = 10
```

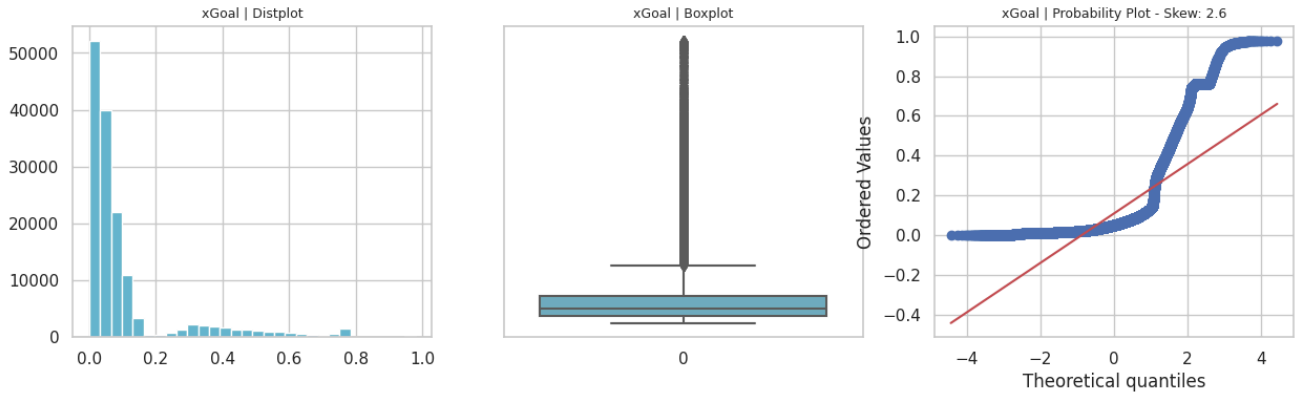
Pair-wise Scatter Plot of all Continuous Variables

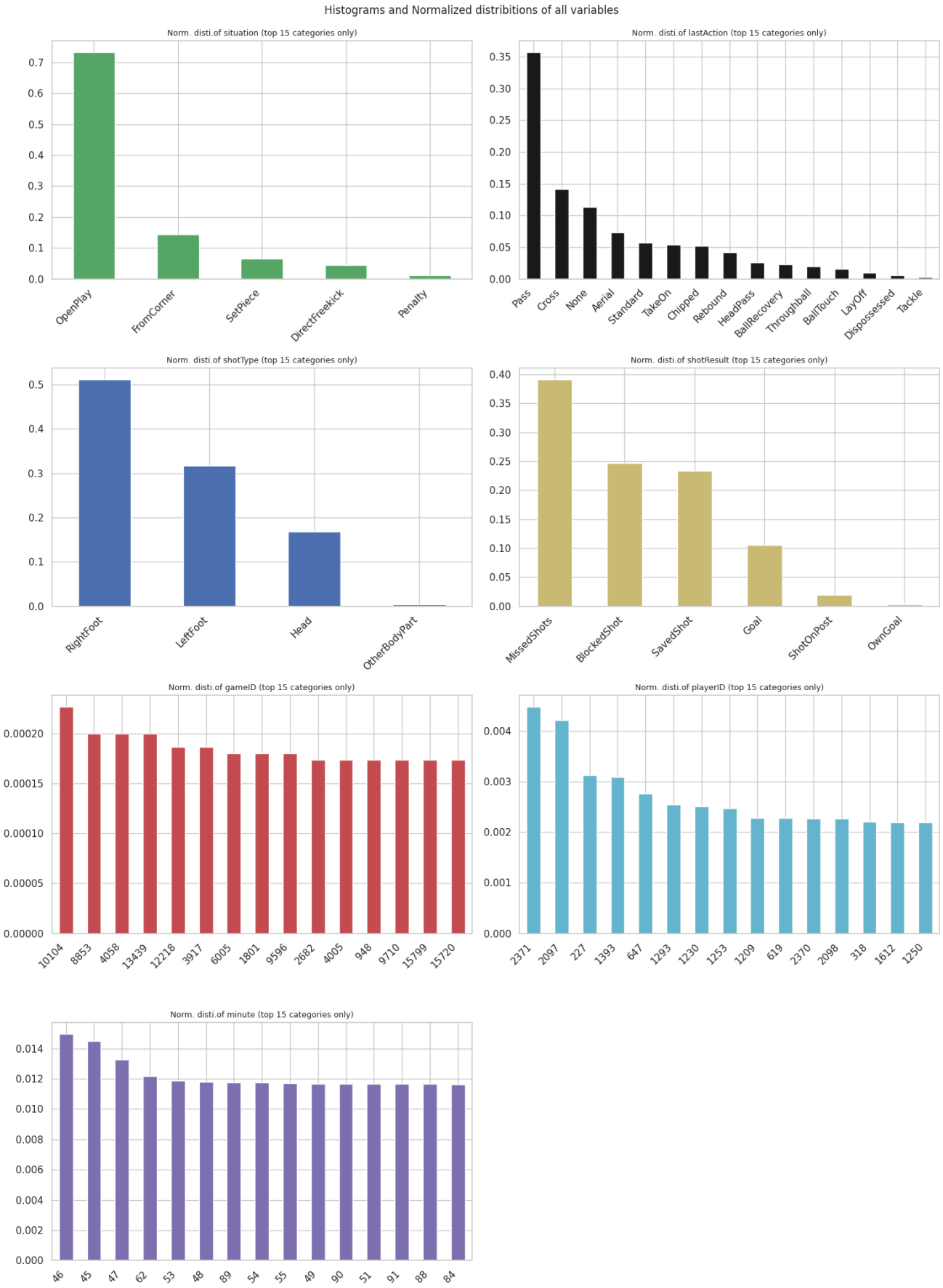



assisterID | Distplot
assisterID | Boxplot
assisterID | Probability Plot - Skew: 1.0
Ordered Values
Theoretical quantiles

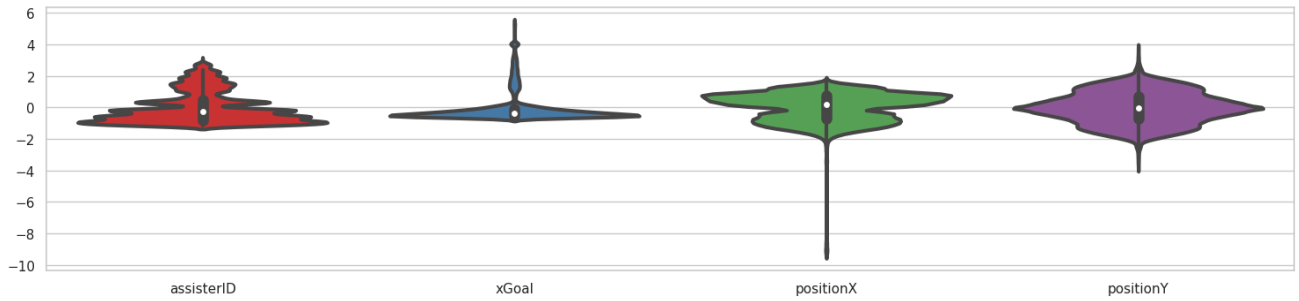

positionY | Distplot
positionY | Boxplot
positionY | Probability Plot - Skew: 0.0
Ordered Values
Theoretical quantiles


positionX | Distplot
positionX | Boxplot
positionX | Probability Plot - Skew: -2.3
Ordered Values
Theoretical quantiles

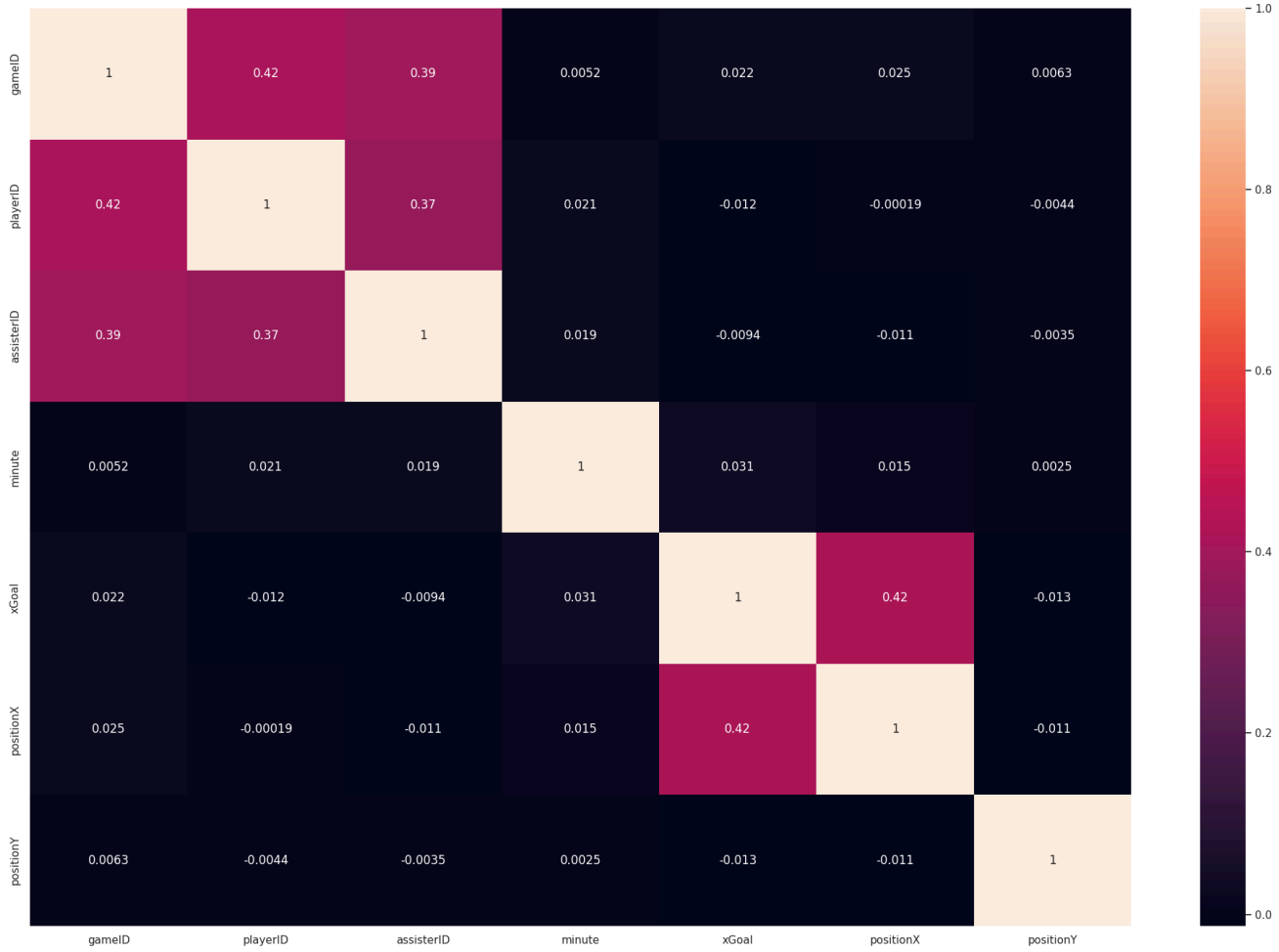

xGoal | Distplot
xGoal | Boxplot
xGoal | Probability Plot - Skew: 2.6
Ordered Values
Theoretical quantiles


Histograms and Normalized distribitions of all variables
Norm. disti.of situation (top 15 categories only)
OpenPlay
FromCorner
SetPiece
DirectFreekick
Penalty
Norm. disti.of lastAction (top 15 categories only)
Pass
Cross
None
Aerial
Standard
TakeOn
Chipped
Rebound
HeadPass
BallRecovery
Throughball
BallTouch
LayOff
Dispossessed
Tackle
Norm. disti.of shotType (top 15 categories only)
RightFoot
LeftFoot
Head
OtherBodyPart
Norm. disti.of shotResult (top 15 categories only)
MissedShots
BlockedShot
SavedShot
Goal
ShotOnPost
OwnGoal
Norm. disti.of gameID (top 15 categories only)
10104
8853
4058
13439
12218
3917
6005
1801
9596
2682
4005
948
9710
15799
15720
Norm. disti.of playerID (top 15 categories only)
2371
2097
227
1393
647
1293
1230
1253
1209
619
2370
2098
318
1612
1250
Norm. disti.of minute (top 15 categories only)
46
45
47
62
53
48
89
54
55
49
90
51
91
88
84

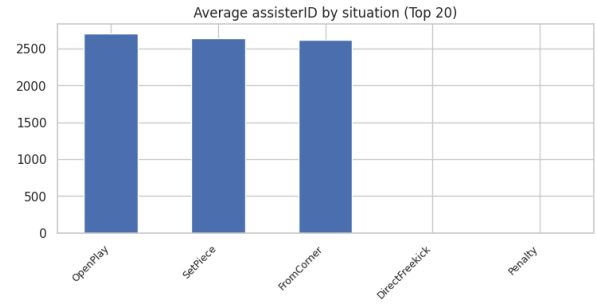
Violin Plot of all Continuous Variables

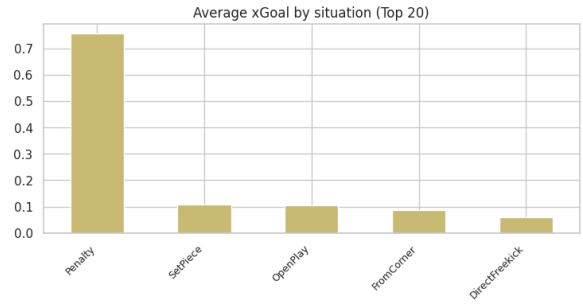


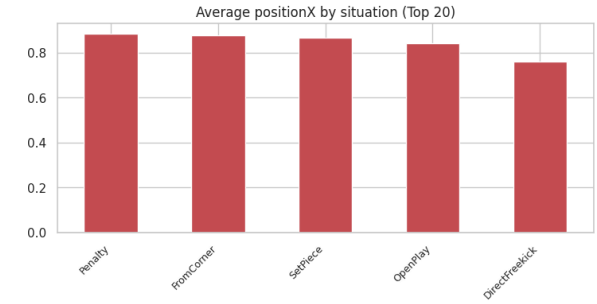
Heatmap of all Numeric Variables including target:

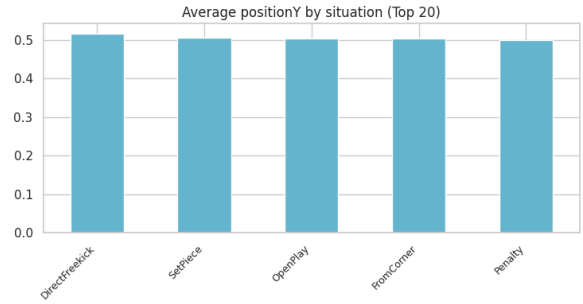


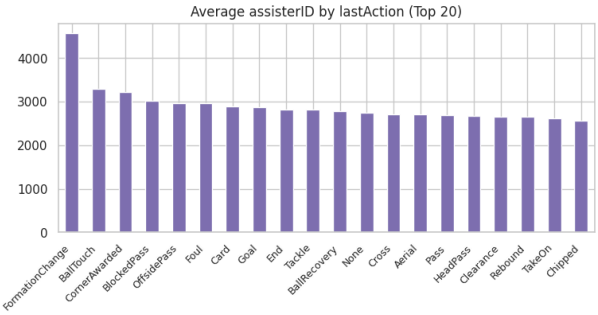
Bar plots for each Continuous by each Categorical variable

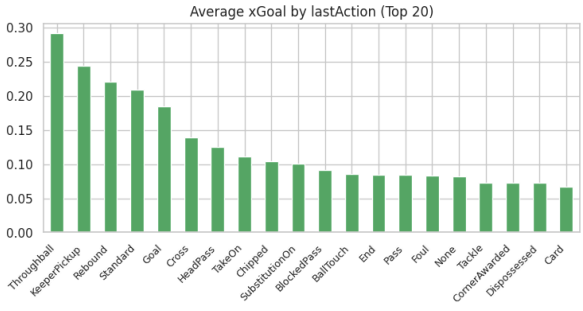


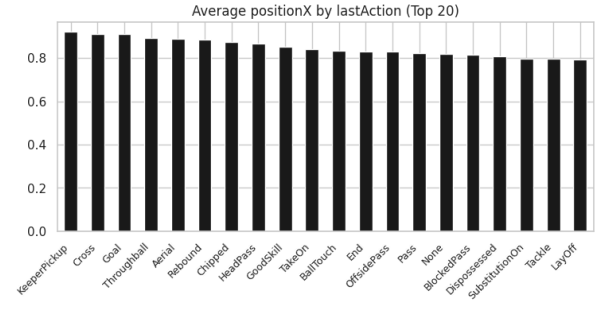


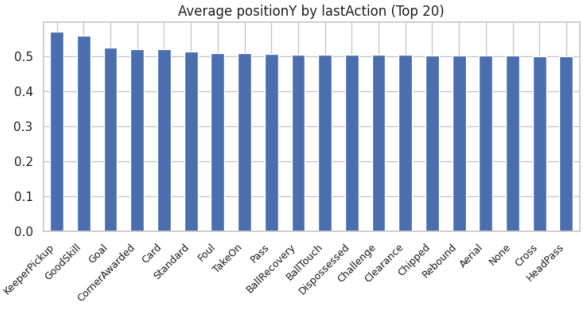


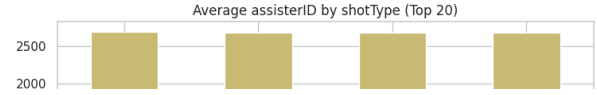


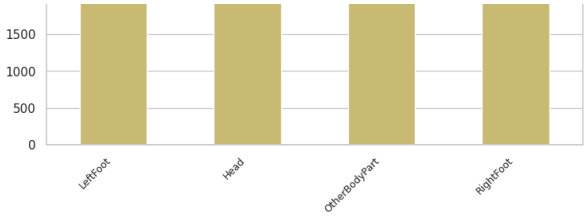


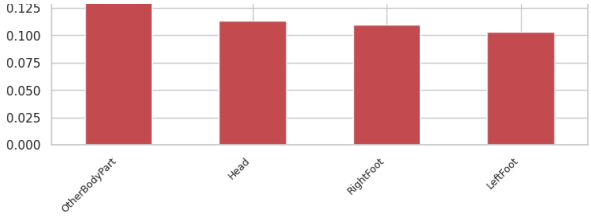


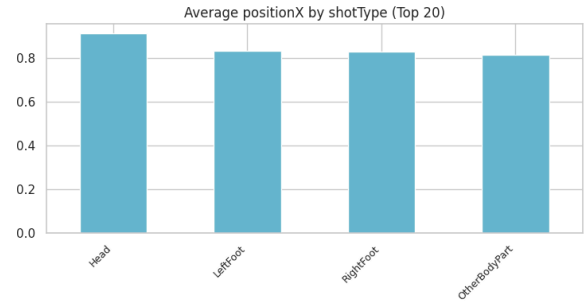


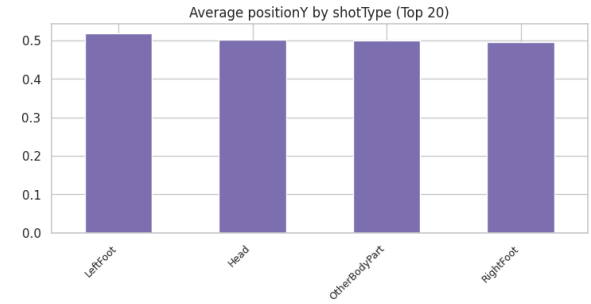


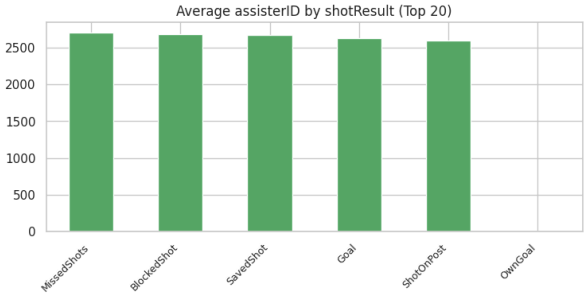


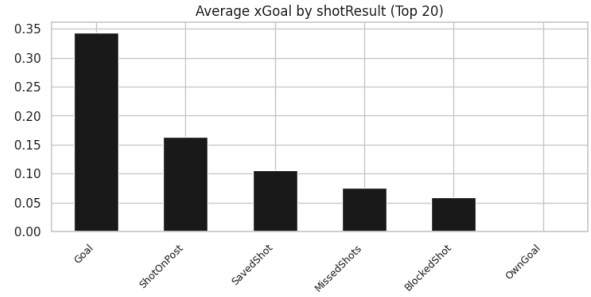

1500
1000
500
0
LeftFoot
Head
OtherBodyPart
RightFoot

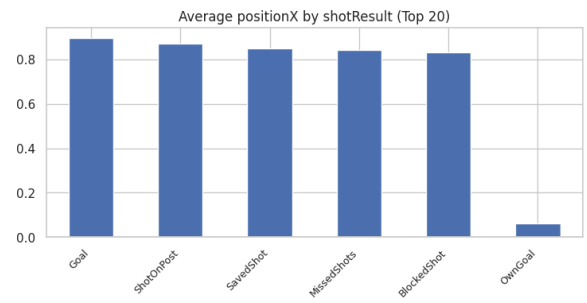

0.125
0.100
0.075
0.050
0.025
0.000
OtherBodyPart
Head
RightFoot
LeftFoot

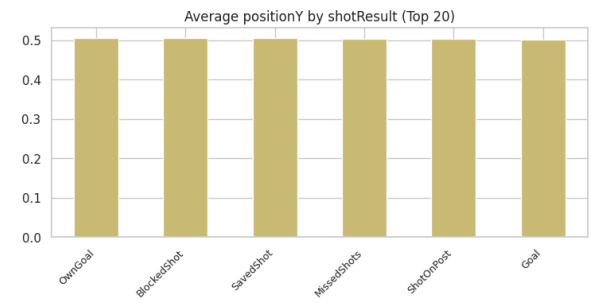

Average positionX by shotType (Top 20)
0.8
0.6
0.4
0.2
0.0
Head
LeftFoot
RightFoot
OtherBodyPart

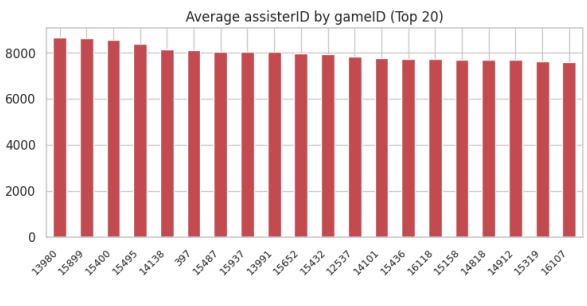

Average positionY by shotType (Top 20)
0.5
0.4
0.3
0.2
0.1
0.0
LeftFoot
Head
OtherBodyPart
RightFoot

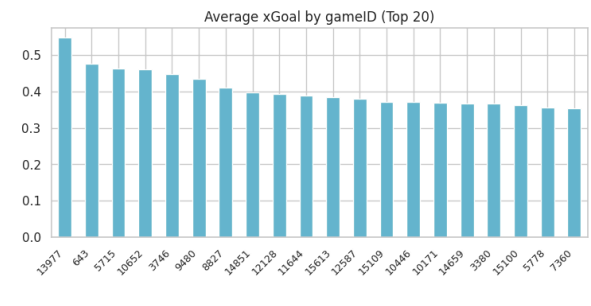

Average assisterID by shotResult (Top 20)
2500
2000
1500
1000
500
0
MissedShots
BlockedShot
SavedShot
Goal
ShotOnPost
OwnGoal

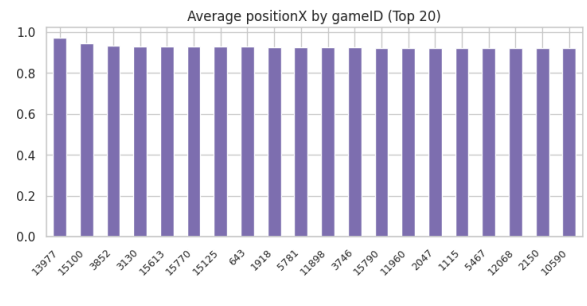

Average xGoal by shotResult (Top 20)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Goal
ShotOnPost
SavedShot
MissedShots
BlockedShot
OwnGoal

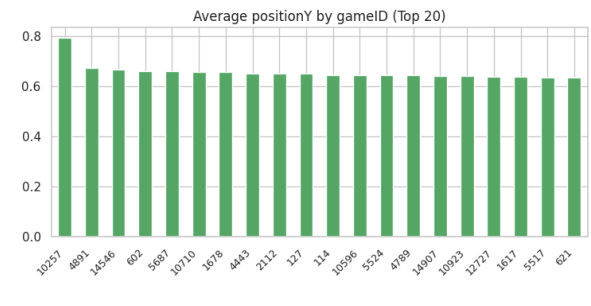

Average positionX by shotResult (Top 20)
0.8
0.6
0.4
0.2
0.0
Goal
ShotOnPost
SavedShot
MissedShots
BlockedShot
OwnGoal

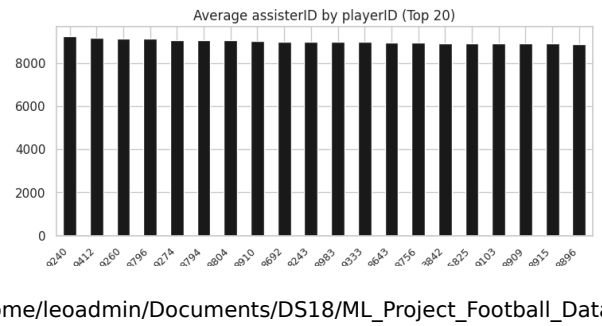

Average positionY by shotResult (Top 20)
0.5
0.4
0.3
0.2
0.1
0.0
OwnGoal
BlockedShot
SavedShot
MissedShots
ShotOnPost
Goal

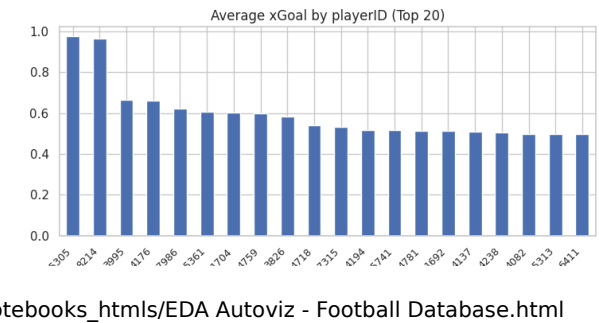

Average assisterID by gameID (Top 20)
8000
6000
4000
2000
0
13960
15899
15400
15495
14138
397
15487
15937
13991
15652
15432
12537
14101
15436
16118
15158
14818
14912
15319
16107

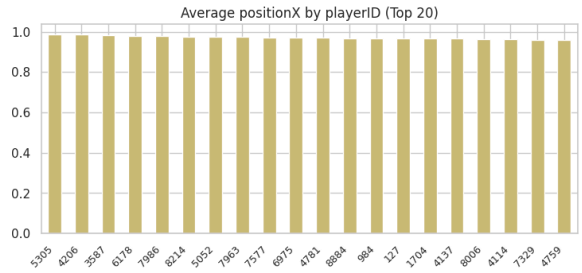

Average xGoal by gameID (Top 20)
0.5
0.4
0.3
0.2
0.1
0.0
13977
643
5715
10652
3746
9480
8827
14851
12128
11644
15613
12587
15109
10446
10171
14659
3380
15100
5778
7360

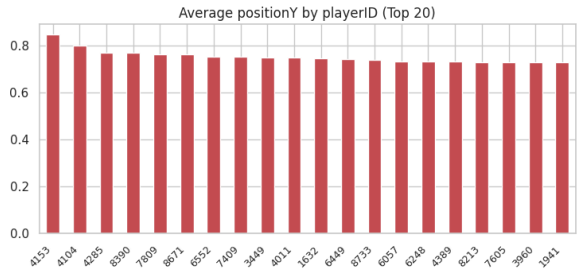

Average positionX by gameID (Top 20)
1.0
0.8
0.6
0.4
0.2
0.0
13977
15100
3852
3130
15613
15770
15125
643
1918
5781
11898
3746
15790
11960
2047
1115
5467
12068
2150
10590

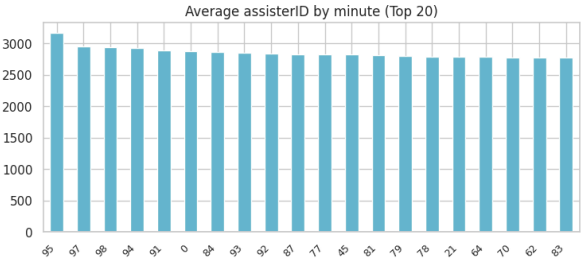

Average positionY by gameID (Top 20)
0.8
0.6
0.4
0.2
0.0
10257
4891
14546
602
5687
10710
1678
4443
2112
127
114
10596
5524
4789
14907
10923
12727
1617
5517
621

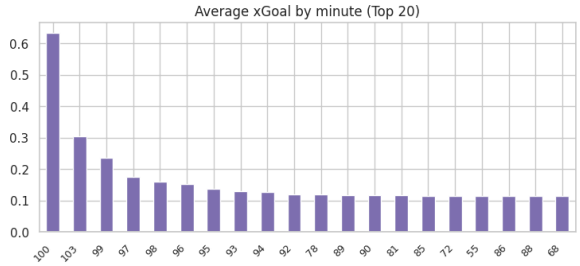

Average assisterID by playerID (Top 20)
8000
6000
4000
2000
0

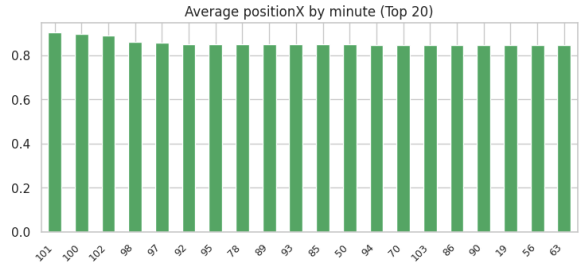

Average xGoal by playerID (Top 20)
1.0
0.8
0.6
0.4
0.2
0.0

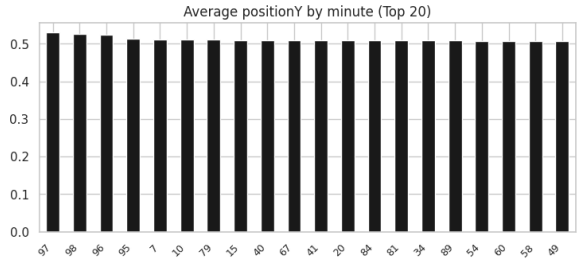

Average positionX by playerID (Top 20)
1.0
0.8
0.6
0.4
0.2
0.0
5305 4206 3587 6178 7986 8214 5052 7963 7577 6975 4781 8884 984 127 1704 4137 8006 4114 7329 4759


Average positionY by playerID (Top 20)
0.8
0.6
0.4
0.2
0.0
4153 4104 4285 8390 7809 8671 6552 7409 3449 4011 1632 6449 8733 6057 6248 4389 8213 7605 3960 1941


Average assisterID by minute (Top 20)
3000
2500
2000
1500
1000
500
0
95 97 98 94 91 0 84 93 92 87 77 45 81 79 78 21 64 70 62 83


Average xGoal by minute (Top 20)
0.6
0.5
0.4
0.3
0.2
0.1
0.0
100 103 99 97 98 96 95 93 94 92 78 89 90 81 85 72 55 86 88 68


Average positionX by minute (Top 20)
0.8
0.6
0.4
0.2
0.0
101 100 102 98 97 92 95 78 89 93 85 50 94 70 103 86 90 19 56 63


Average positionY by minute (Top 20)
0.5
0.4
0.3
0.2
0.1
0.0
97 98 96 95 7 10 79 15 40 67 41 20 84 81 34 89 54 60 58 49

```
All Plots done
Time to run AutoViz = 16 seconds

 ###################### AUTO VISUALIZATION Completed ######################
=== Completed AutoViz on player_shots ===

Skipping teamstats...
=== Running AutoViz on df_combined ===
Shape of your Data Set loaded: (12680, 39)
################################################################################
#####
######################## C L A S S I F Y I N G  V A R I A B L E S  ##############
#####
################################################################################
#####
Classifying variables in data set...
    Number of Numeric Columns =  6
    Number of Integer-Categorical Columns =  21
    Number of String-Categorical Columns =  3
    Number of Factor-Categorical Columns =  0
    Number of String-Boolean Columns =  0
    Number of Numeric-Boolean Columns =  0
    Number of Discrete String Columns =  0
    Number of NLP String Columns =  3
    Number of Date Time Columns =  3
    Number of ID Columns =  1
    Number of Columns to Delete =  2
    39 Predictors classified...
        3 variable(s) removed since they were ID or low-information variables
        List of variables removed: ['gameID', 'home_location', 'away_location']
6 numeric variables in data exceeds limit, taking top 30 variables
    List of variables selected: ['home_xGoals', 'home_ppda', 'away_xGoals', 'away_
ppda', 'home_yellowCards', 'away_yellowCards']
   Total columns > 30, too numerous to print.
To fix these data quality issues in the dataset, import FixDQ from autoviz...
    All variables classified into correct types.
```

	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ Issue
gameID	int64	0.000000	100	81.000000	16135.000000	Possible ID column: drop before modeling step.
leagueID	int64	0.000000	0	1.000000	5.000000	No issue
season	int64	0.000000	0	2014.000000	2020.000000	Possible date-time colum: transform before modeling step.
date	object	0.000000	53			No issue
homeTeamID	int64	0.000000	1	71.000000	262.000000	Column has 55 outliers greater than upper bound (256.00) or lower than lower bound(8.00). Cap them or remove them.
awayTeamID	int64	0.000000	1	71.000000	262.000000	Column has 55 outliers greater than upper bound (256.00) or lower than lower bound(8.00). Cap them or remove them.
homeGoals	int64	0.000000	0	0.000000	10.000000	Column has 981 outliers greater than upper bound (3.50) or lower than lower bound(-0.50). Cap them or remove them.
awayGoals	int64	0.000000	0	0.000000	9.000000	Column has 44 outliers greater than upper bound (5.00) or lower than lower bound(-3.00). Cap them or remove them.
homeGoalsHalfTime	int64	0.000000	0	0.000000	6.000000	Column has 403 outliers greater than upper bound (2.50) or lower than lower bound(-1.50). Cap them or remove them.
awayGoalsHalfTime	int64	0.000000	0	0.000000	5.000000	Column has 225 outliers greater than upper bound (2.50) or lower than lower bound(-1.50). Cap them or remove them.

	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ Issue
home_season	int64	0.000000	0	2014.000000	2020.000000	Possible date-time colum: transform before modeling step.
home_date	object	0.000000	53			No issue
home_location	object	0.000000	0			Possible Zero-variance or low information colum: drop before modeling step.
home_goals	int64	0.000000	0	0.000000	10.000000	Column has 981 outliers greater than upper bound (3.50) or lower than lower bound(-0.50). Cap them or remove them., Column has a high correlation with ['homeGoals']. Consider dropping one of them.
home_xGoals	float64	0.000000	NA	0.000000	6.630490	Column has 276 outliers greater than upper bound (3.76) or lower than lower bound(-0.92). Cap them or remove them.
home_shots	int64	0.000000	0	0.000000	47.000000	Column has 167 outliers greater than upper bound (27.50) or lower than lower bound(-0.50). Cap them or remove them.
home_shotsOnTarget	int64	0.000000	0	0.000000	18.000000	Column has 351 outliers greater than upper bound (10.50) or lower than lower bound(-1.50). Cap them or remove them.
home_deep	int64	0.000000	0	0.000000	42.000000	Column has 227 outliers greater than upper bound (18.00) or lower than lower bound(-6.00). Cap them or remove them.
home_ppda	float64	0.000000	NA	1.897400	97.333300	Column has 540 outliers greater than upper bound (21.31) or lower than lower bound(-1.88). Cap

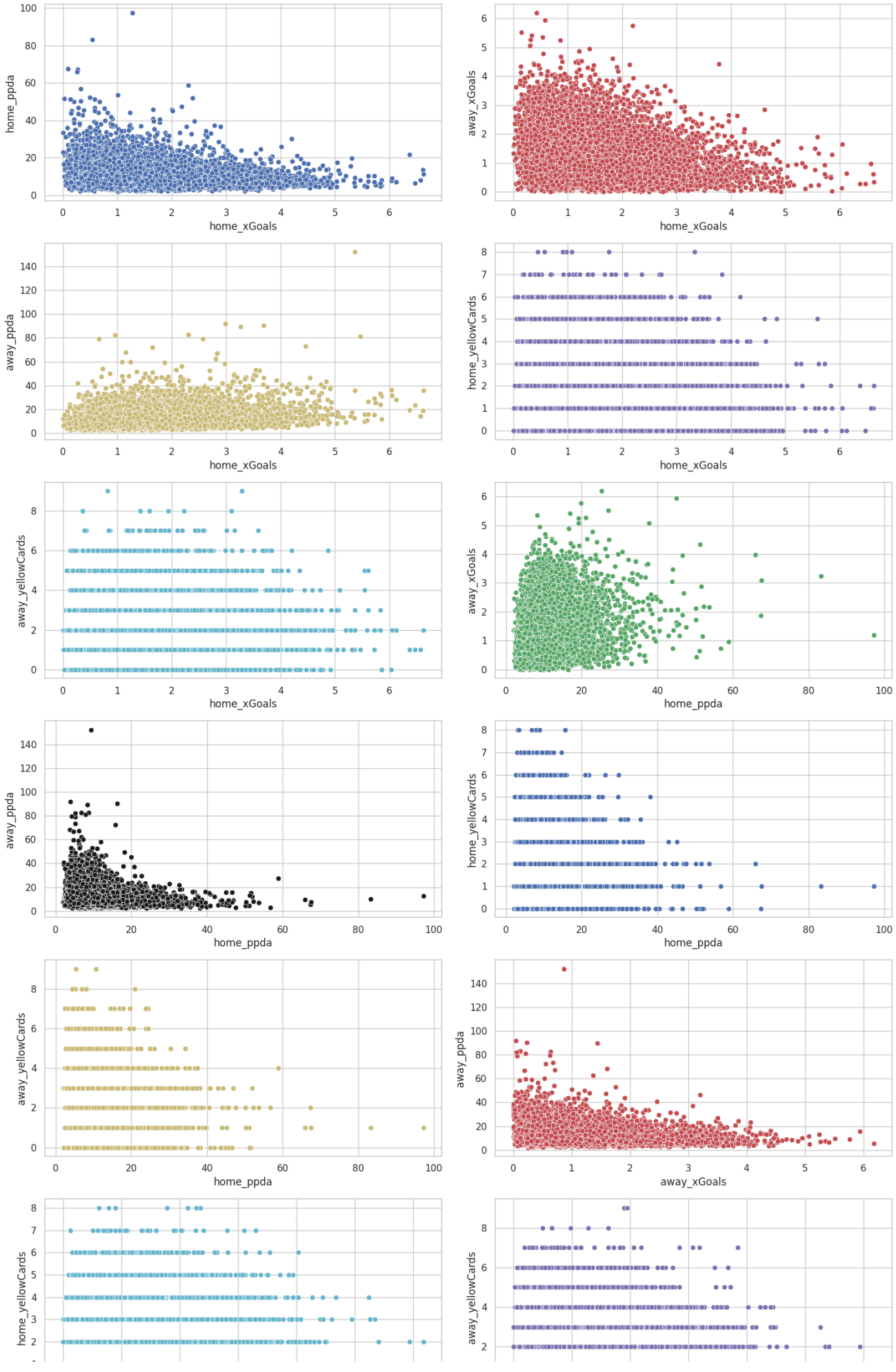
	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ Issue
						them or remove them.
home_fouls	int64	0.000000	0	0.000000	33.000000	Column has 241 outliers greater than upper bound (22.50) or lower than lower bound(2.50). Cap them or remove them.
home_corners	int64	0.000000	0	0.000000	20.000000	Column has 143 outliers greater than upper bound (13.00) or lower than lower bound(-3.00). Cap them or remove them.
home_yellowCards	float64	0.007886	NA	0.000000	8.000000	1 missing values. Impute them with mean, median, mode, or a constant value such as 123., Column has 37 outliers greater than upper bound (6.00) or lower than lower bound(-2.00). Cap them or remove them.
home_redCards	int64	0.000000	0	0.000000	3.000000	Column has 1078 outliers greater than upper bound (0.00) or lower than lower bound(0.00). Cap them or remove them.
home_result	object	0.000000	0			No issue
away_season	int64	0.000000	0	2014.000000	2020.000000	Possible date-time colum: transform before modeling step.
away_date	object	0.000000	53			No issue
away_location	object	0.000000	0			Possible Zero-variance or low information colum: drop before modeling step.
away_goals	int64	0.000000	0	0.000000	9.000000	Column has 44 outliers greater than upper bound (5.00) or lower than lower bound(-3.00). Cap them or remove them., Column has a high correlation with ['awayGoals'].

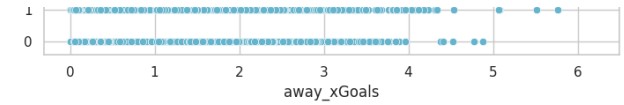
	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ Issue
						Consider dropping one of them.
away_xGoals	float64	0.000000	NA	0.000000	6.186960	Column has 312 outliers greater than upper bound (3.10) or lower than lower bound(-0.91). Cap them or remove them.
away_shots	int64	0.000000	0	0.000000	39.000000	Column has 161 outliers greater than upper bound (23.00) or lower than lower bound(-1.00). Cap them or remove them.
away_shotsOnTarget	int64	0.000000	0	0.000000	15.000000	Column has 233 outliers greater than upper bound (9.50) or lower than lower bound(-2.50). Cap them or remove them.
away_deep	int64	0.000000	0	0.000000	28.000000	Column has 423 outliers greater than upper bound (13.00) or lower than lower bound(-3.00). Cap them or remove them.
away_ppda	float64	0.000000	NA	2.122000	152.000000	Column has 606 outliers greater than upper bound (24.25) or lower than lower bound(-2.58). Cap them or remove them.
away_fouls	int64	0.000000	0	0.000000	32.000000	Column has 81 outliers greater than upper bound (25.00) or lower than lower bound(1.00). Cap them or remove them.
away_corners	int64	0.000000	0	0.000000	19.000000	Column has 283 outliers greater than upper bound (10.50) or lower than lower bound(-1.50). Cap them or remove them.
away_yellowCards	float64	0.000000	NA	0.000000	9.000000	Column has 43 outliers greater than upper bound (6.00) or lower than lower

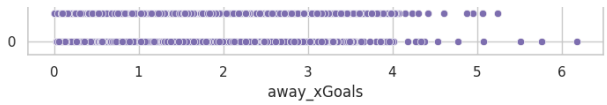
	Data Type	Missing Values%	Unique Values%	Minimum Value	Maximum Value	DQ Issue
						bound(-2.00). Cap them or remove them.
away_redCards	int64	0.000000	0	0.000000	3.000000	Column has 1396 outliers greater than upper bound (0.00) or lower than lower bound(0.00). Cap them or remove them.
away_result	object	0.000000	0			No issue
gameresult	object	0.000000	0			No issue

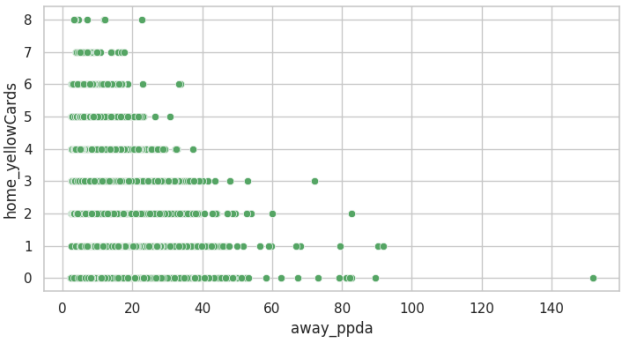
```
Number of All Scatter Plots = 21
```

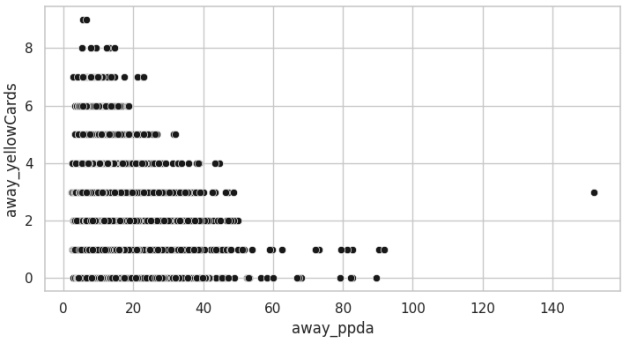
Pair-wise Scatter Plot of all Continuous Variables

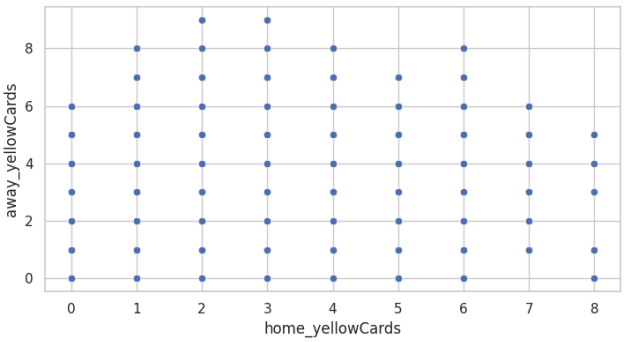


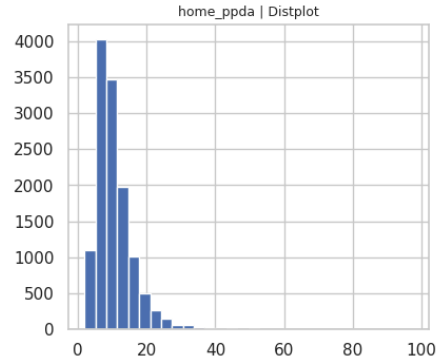

away_xGoals

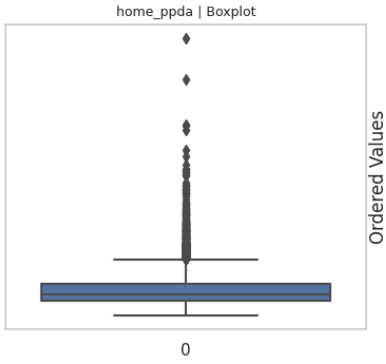

away_xGoals

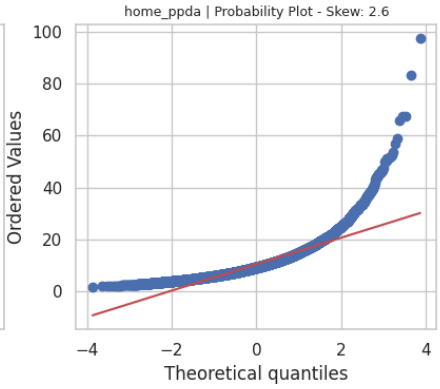

home_yellowCards
away_ppda

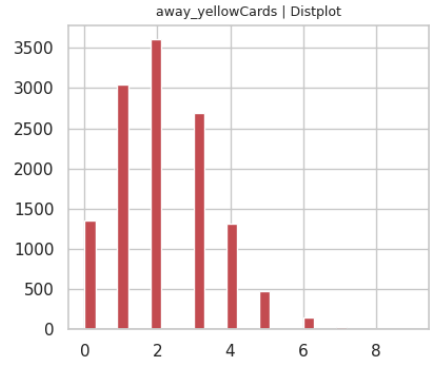

away_yellowCards
away_ppda

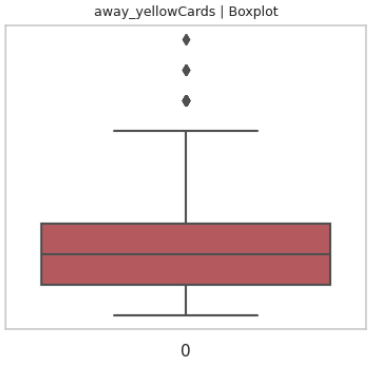

away_yellowCards
home_yellowCards

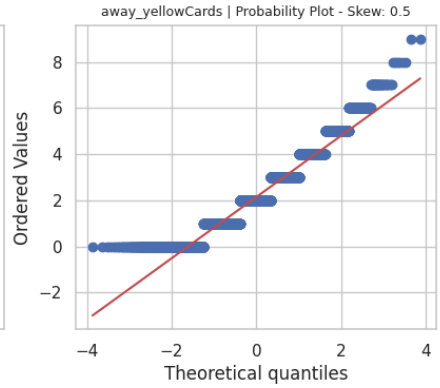

home_ppda | Distplot

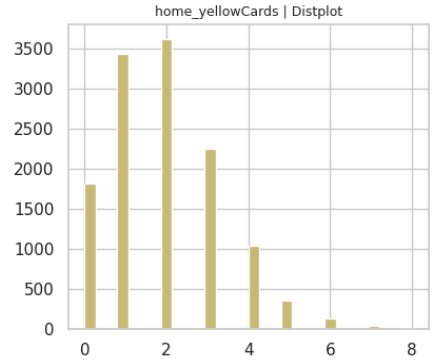

home_ppda | Boxplot

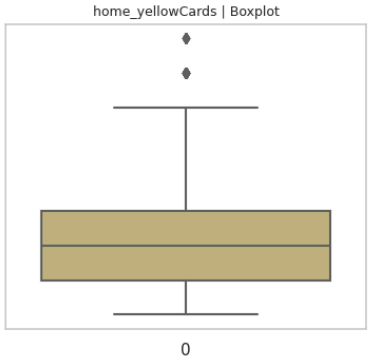

home_ppda | Probability Plot - Skew: 2.6
Ordered Values
Theoretical quantiles

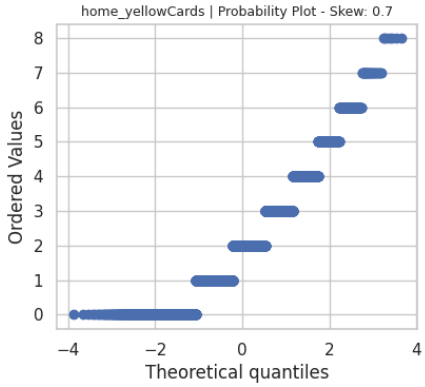

away_yellowCards | Distplot

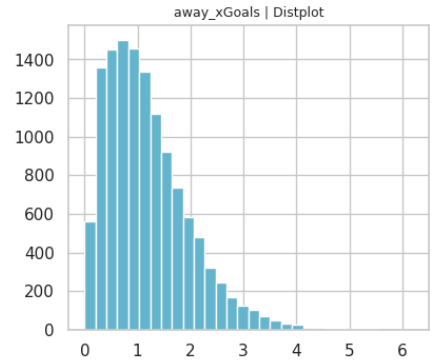

away_yellowCards | Boxplot

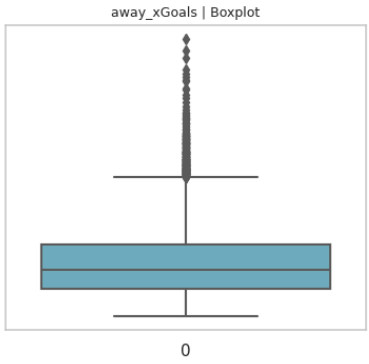

away_yellowCards | Probability Plot - Skew: 0.5
Ordered Values
Theoretical quantiles

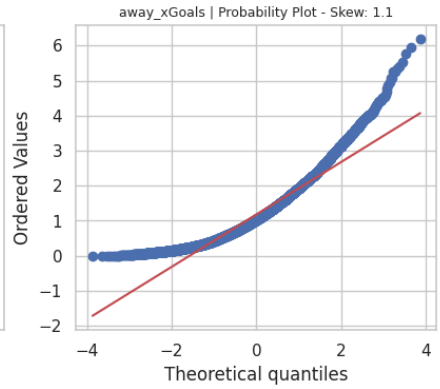

home_yellowCards | Distplot

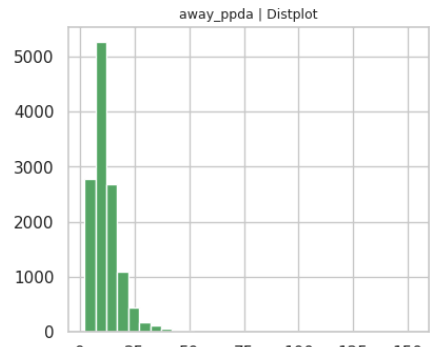

home_yellowCards | Boxplot

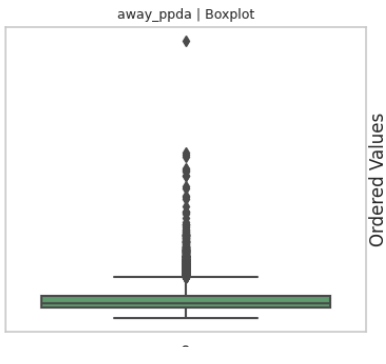

home_yellowCards | Probability Plot - Skew: 0.7
Ordered Values
Theoretical quantiles

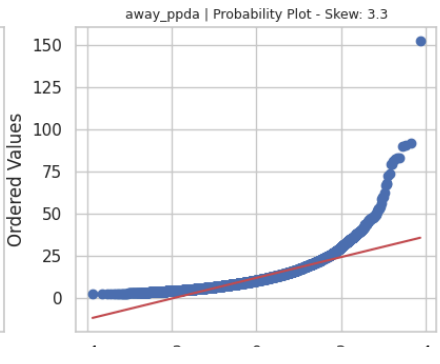

away_xGoals | Distplot

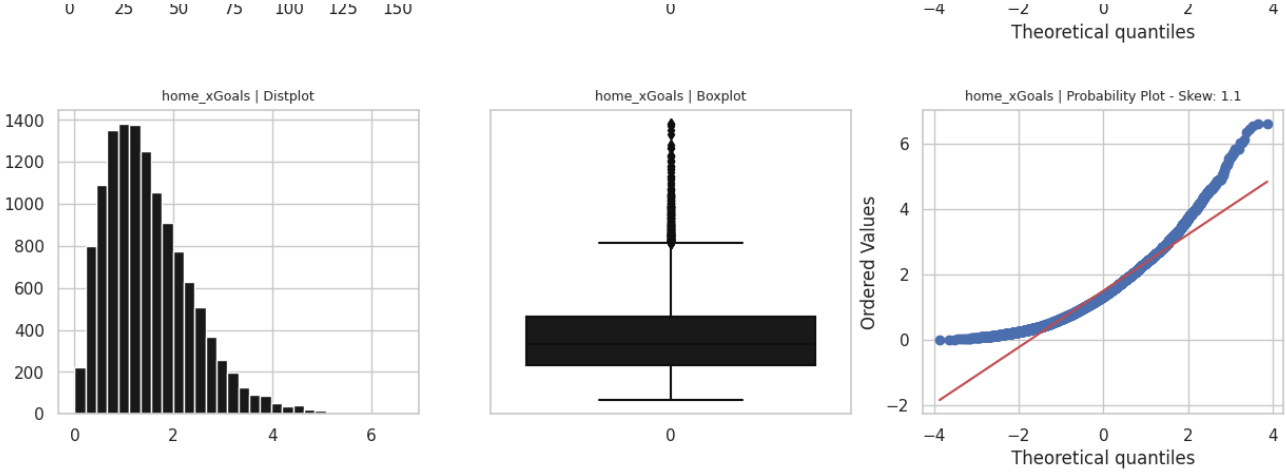

away_xGoals | Boxplot


away_xGoals | Probability Plot - Skew: 1.1
Ordered Values
Theoretical quantiles

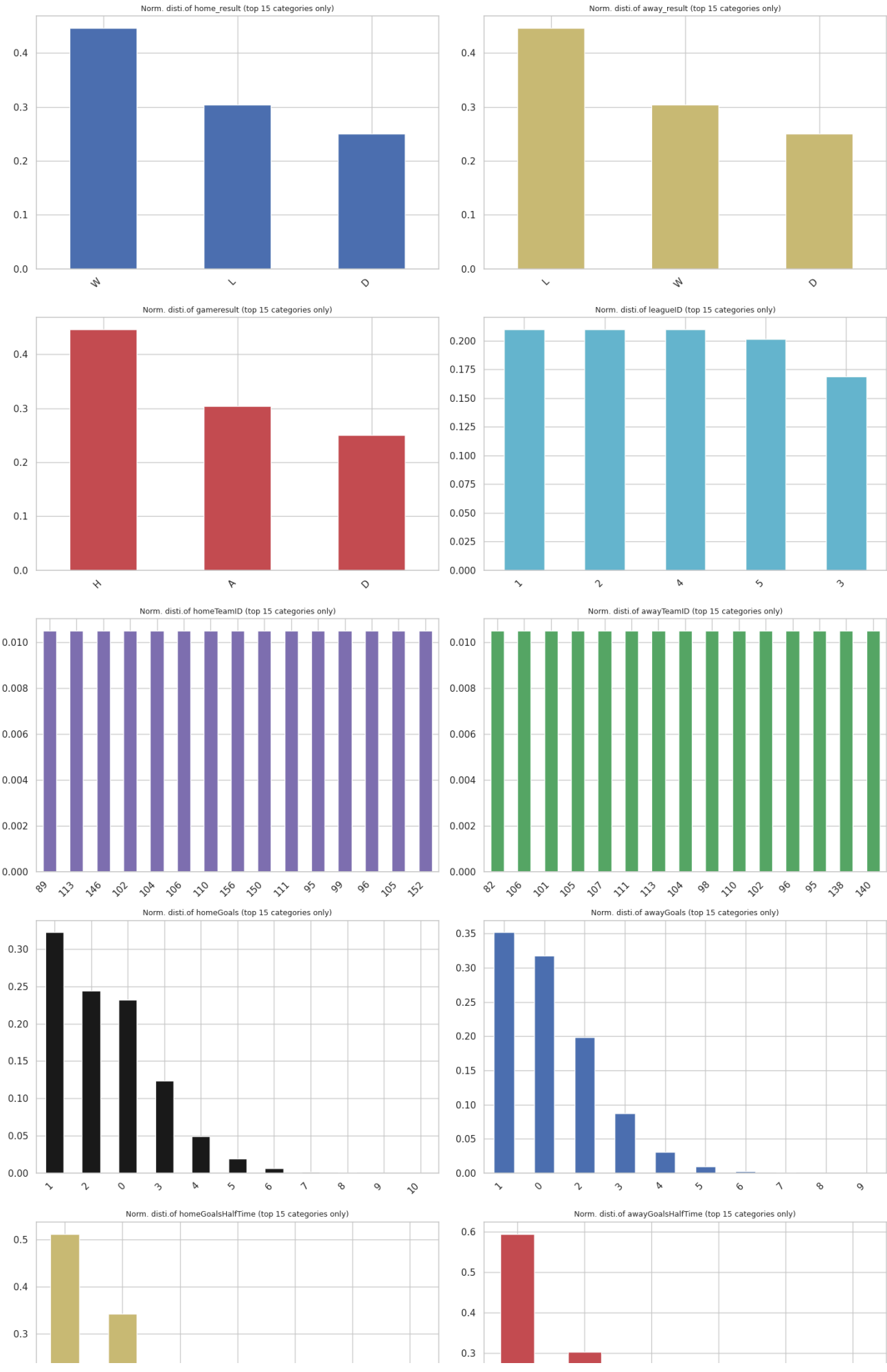

away_ppda | Distplot

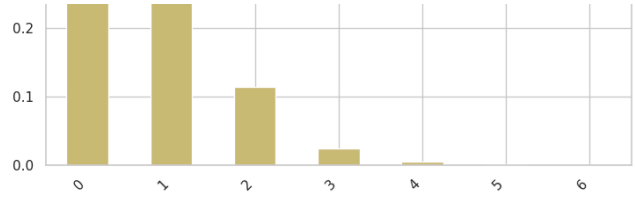

away_ppda | Boxplot

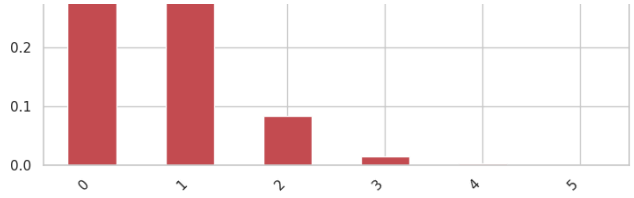

away_ppda | Probability Plot - Skew: 3.3
Ordered Values

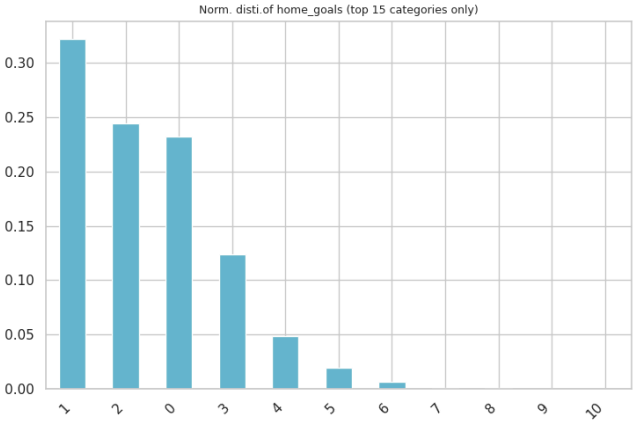

0 25 50 75 100 125 150
0
−4 −2 0 2 4
Theoretical quantiles
home_xGoals | Distplot
1400
1200
1000
800
600
400
200
0
0 2 4 6
home_xGoals | Boxplot
0
home_xGoals | Probability Plot - Skew: 1.1
Ordered Values
6
4
2
0
−2
−4 −2 0 2 4
Theoretical quantiles

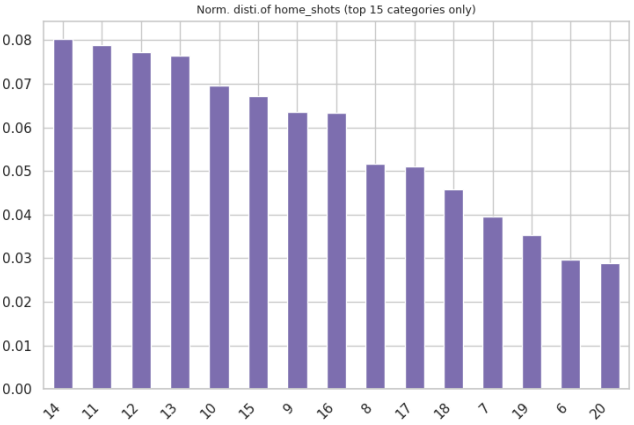
Histograms and Normalized distribitions of all variables

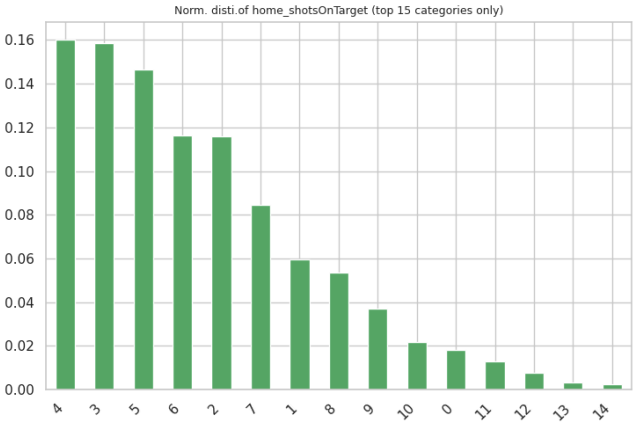


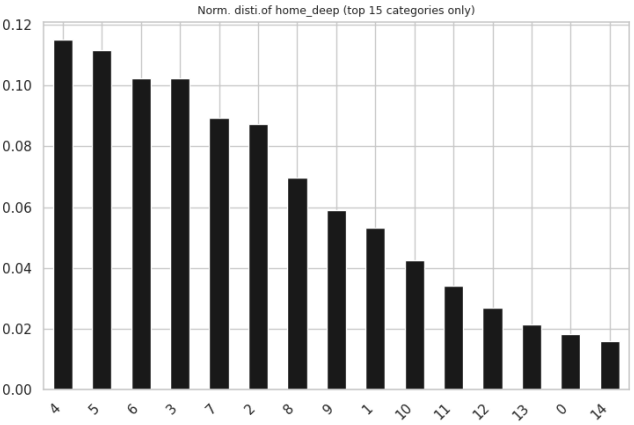


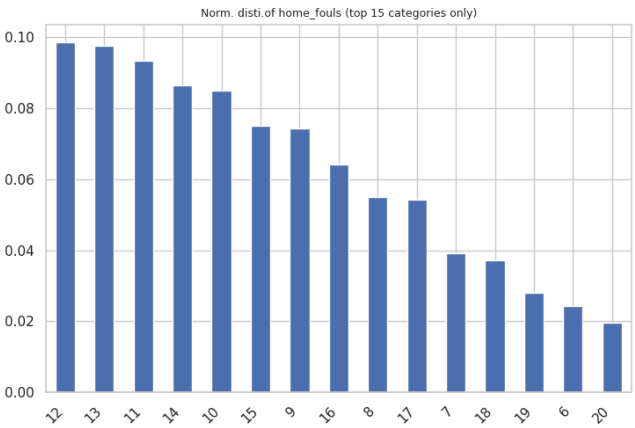


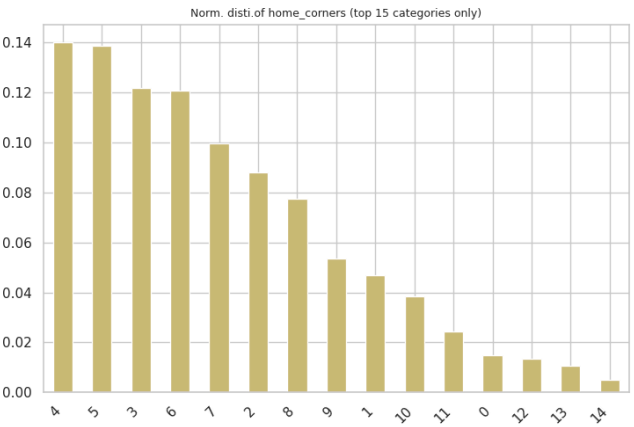

Norm. disti.of home_goals (top 15 categories only)

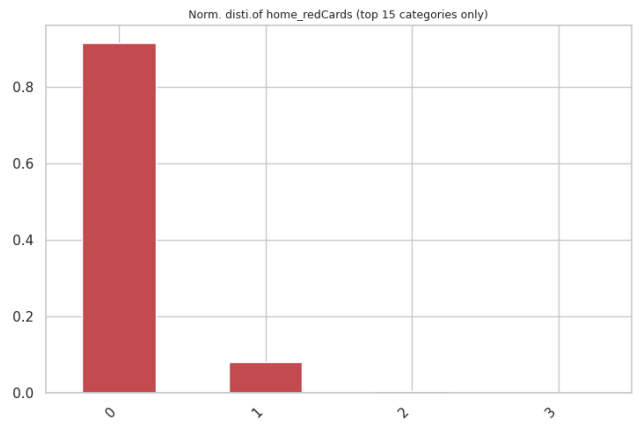

Norm. disti.of home_shots (top 15 categories only)

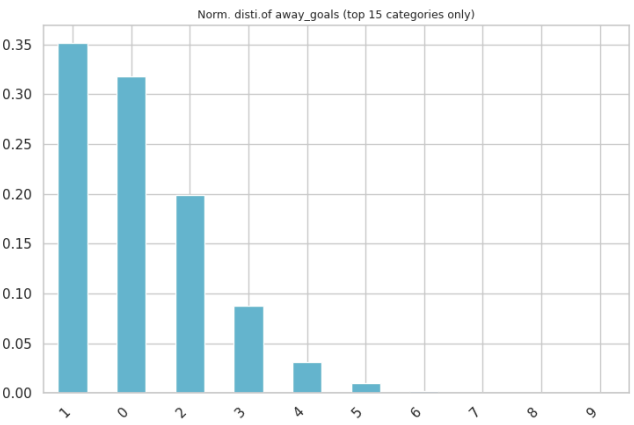

Norm. disti.of home_shotsOnTarget (top 15 categories only)


Norm. disti.of home_deep (top 15 categories only)

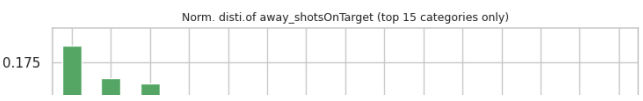

Norm. disti.of home_fouls (top 15 categories only)

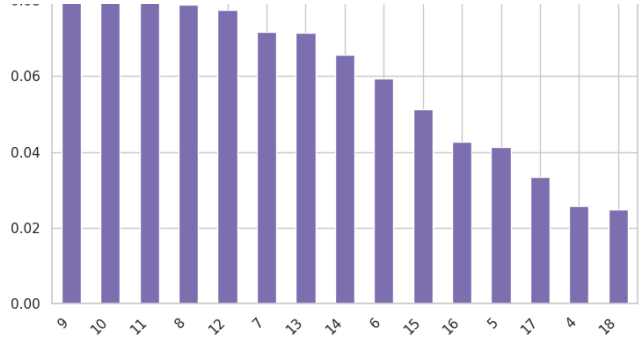

Norm. disti.of home_corners (top 15 categories only)

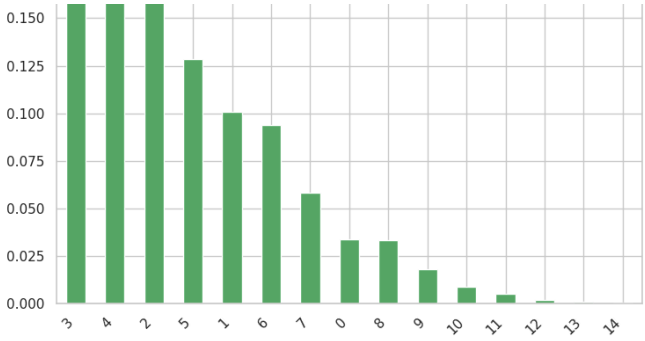

Norm. disti.of home_redCards (top 15 categories only)

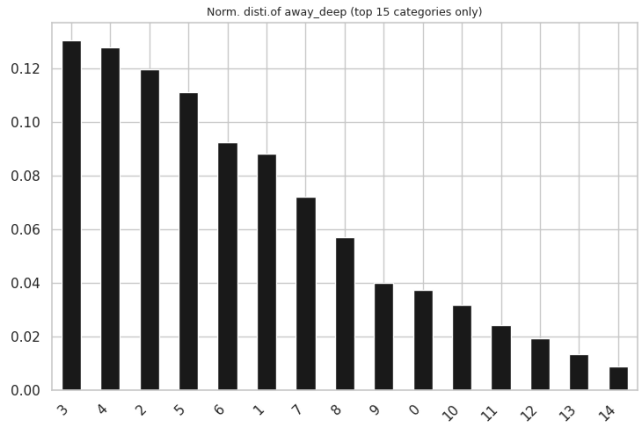

Norm. disti.of away_goals (top 15 categories only)

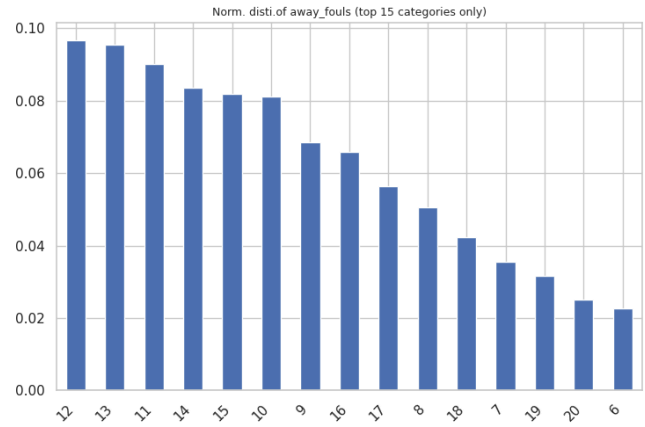

Norm. disti.of away_shots (top 15 categories only)

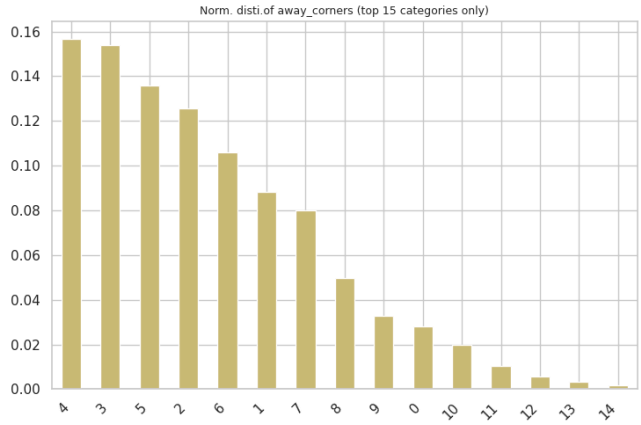

Norm. disti.of away_shotsOnTarget (top 15 categories only)

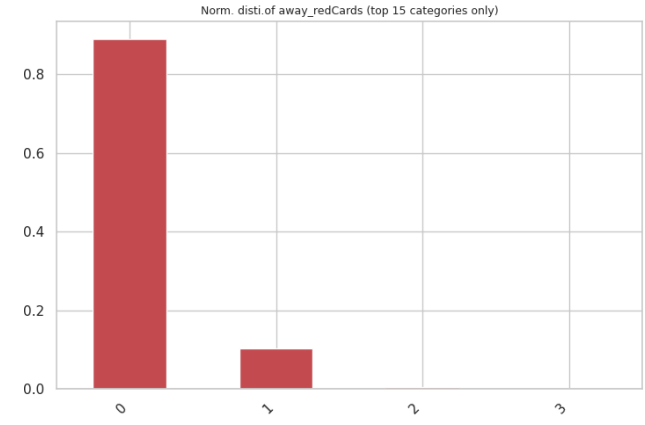


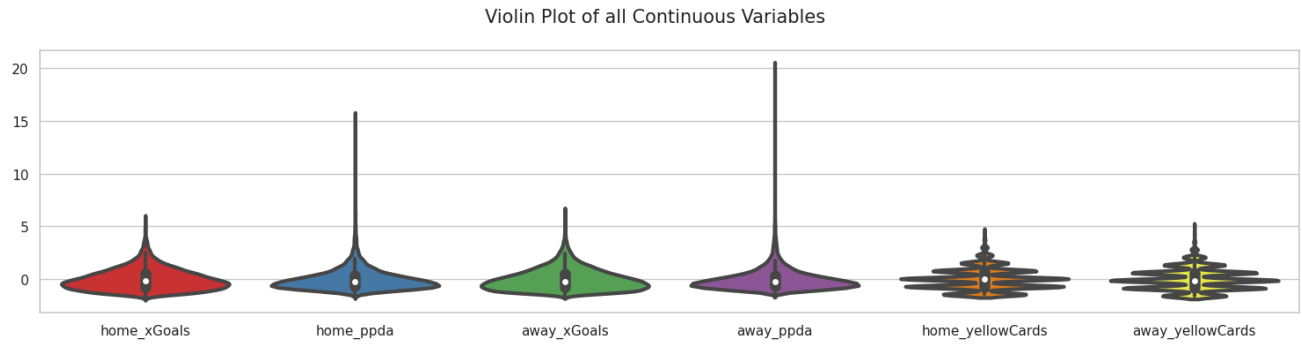


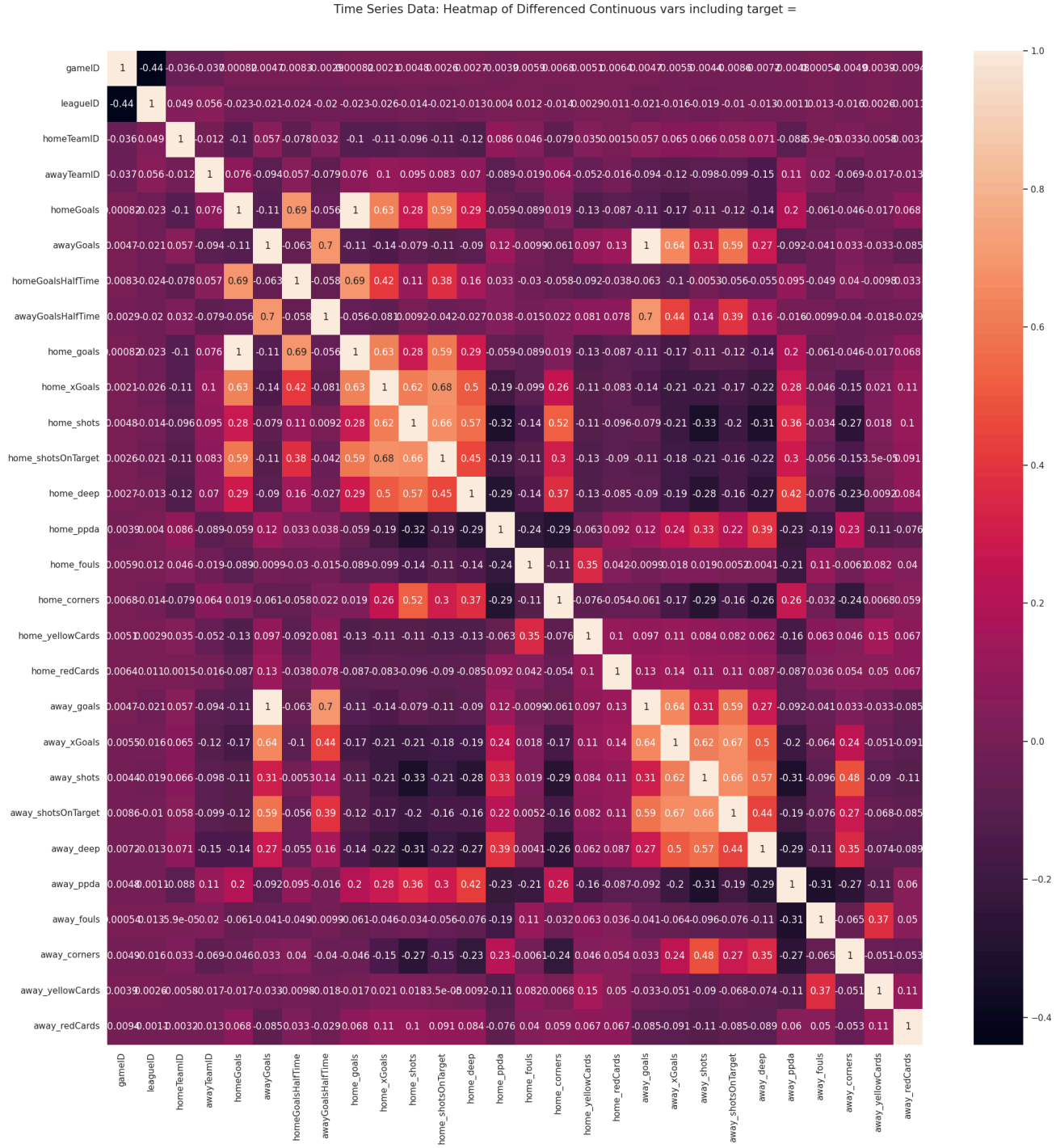

Norm. disti.of away_deep (top 15 categories only)


Norm. disti.of away_fouls (top 15 categories only)

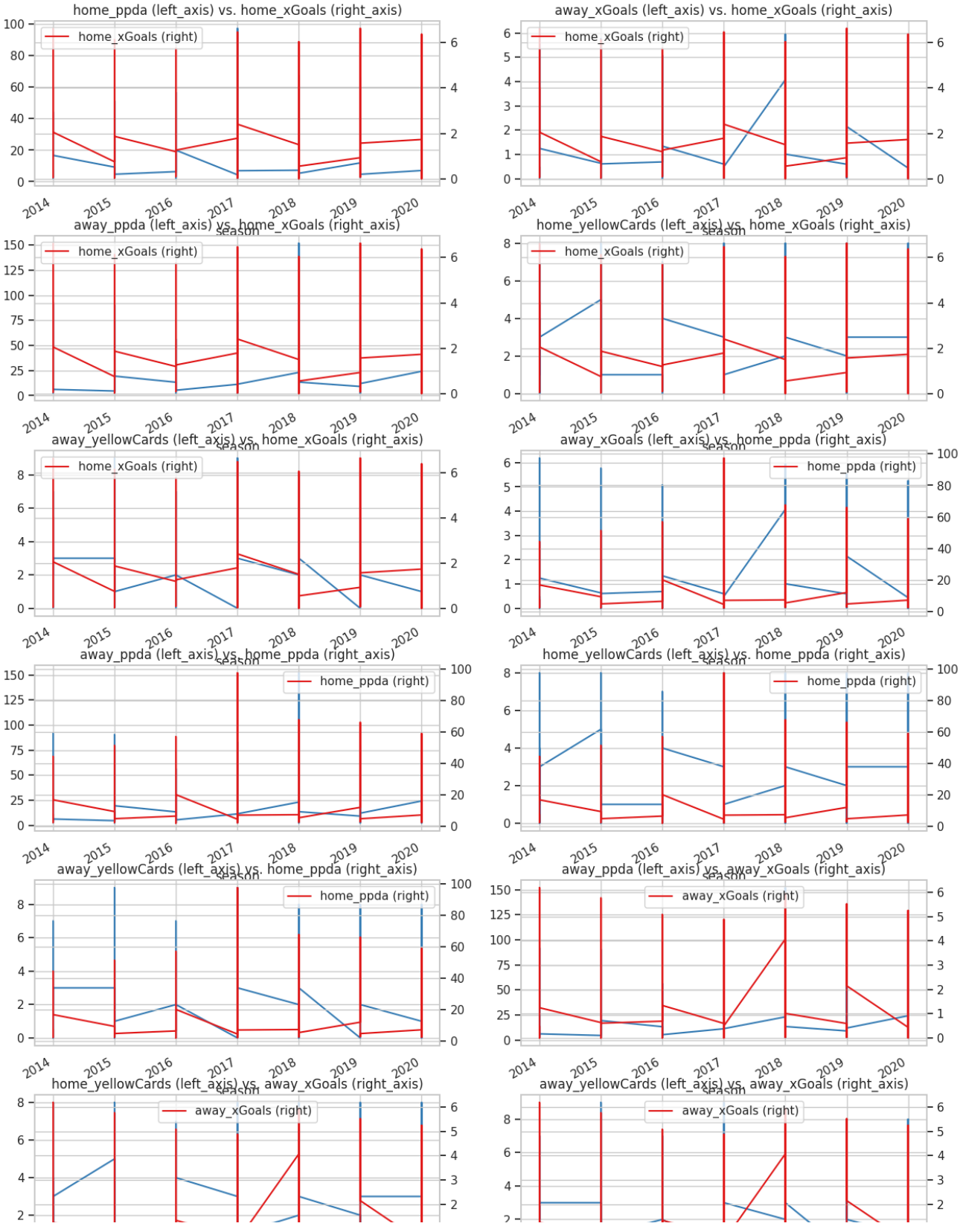

Norm. disti.of away_corners (top 15 categories only)

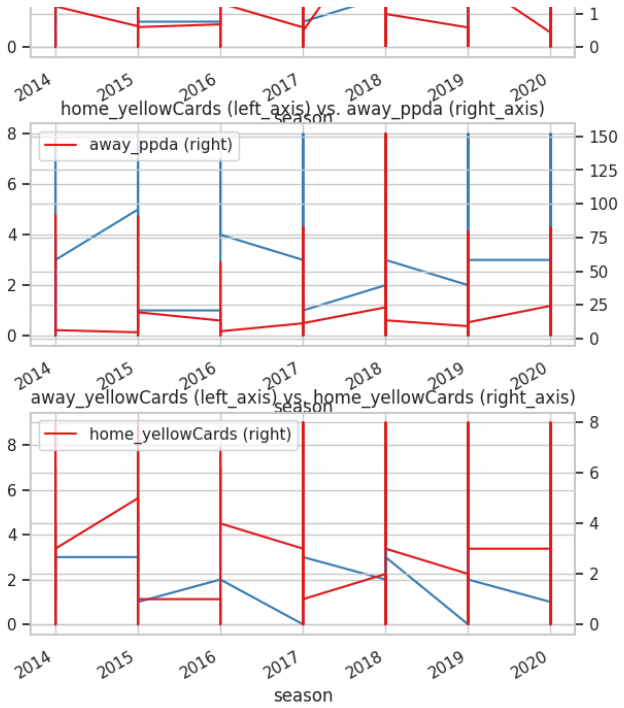

Norm. disti.of away_redCards (top 15 categories only)

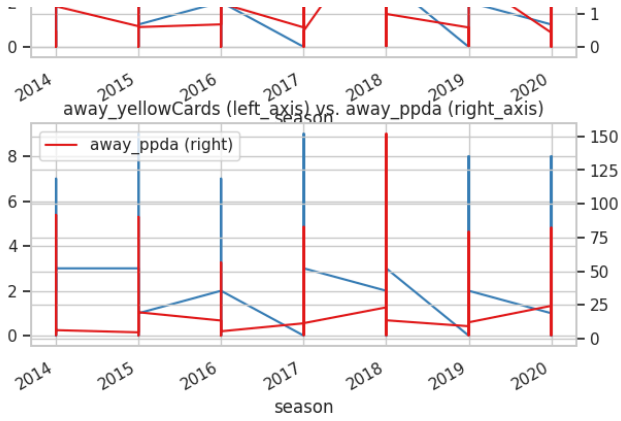

Violin Plot of all Continuous Variables
home_xGoals
home_ppda
away_xGoals
away_ppda
home_yellowCards
away_yellowCards


Time Series Data: Heatmap of Differenced Continuous vars including target =
gameID
leagueID
homeTeamID
awayTeamID
homeGoals
awayGoals
homeGoalsHalfTime
awayGoalsHalfTime
home_goals
home_xGoals
home_shots
home_shotsOnTarget
home_deep
home_ppda
home_fouls
home_corners
home_yellowCards
home_redCards
away_goals
away_xGoals
away_shots
away_shotsOnTarget
away_deep
away_ppda
away_fouls
away_corners
away_yellowCards
away_redCards
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4

Time Series Plot by season: Pairwise Continuous Variables

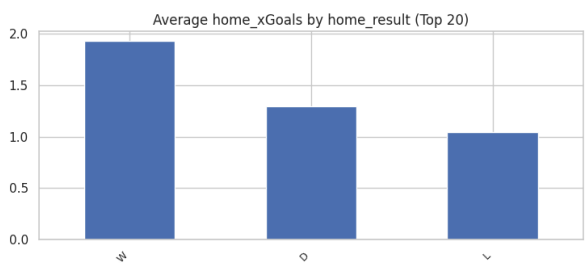


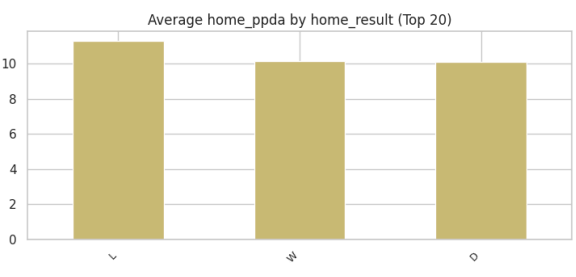

home_yellowCards (left_axis) vs. away_ppda (right_axis)
away_ppda (right)
season

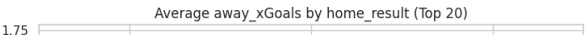

away_yellowCards (left_axis) vs. away_ppda (right_axis)
away_ppda (right)
season

away_yellowCards (left_axis) vs. home_yellowCards (right_axis)
home_yellowCards (right)
season

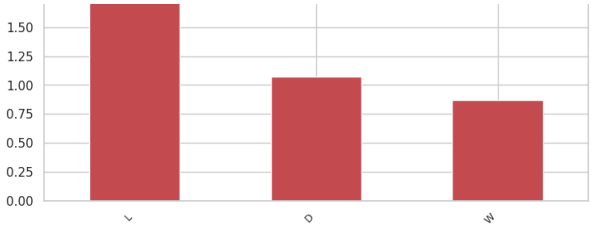
Bar plots for each Continuous by each Categorical variable

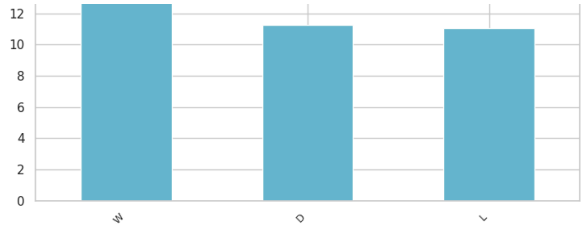

Average home_xGoals by home_result (Top 20)
2.0
1.5
1.0
0.5
0.0
W
D
L

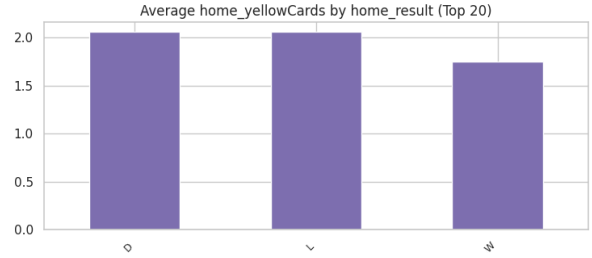

Average home_ppda by home_result (Top 20)
10
8
6
4
2
0
L
W
D

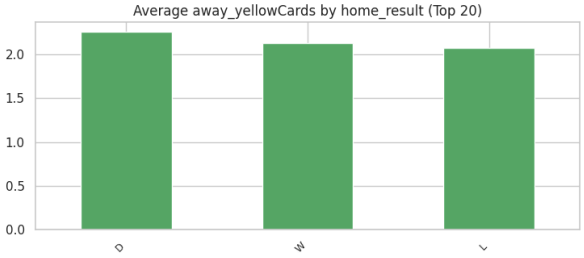

Average away_xGoals by home_result (Top 20)
1.75

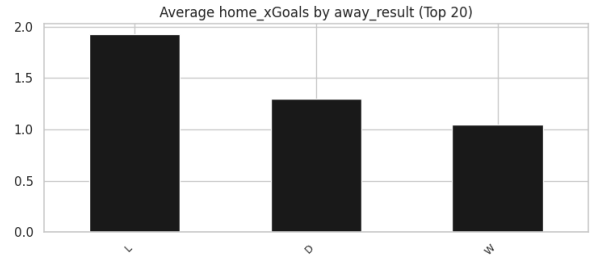

Average away_ppda by home_result (Top 20)

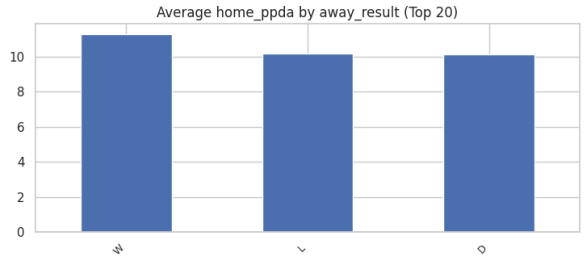

1.50
1.25
1.00
0.75
0.50
0.25
0.00
L
D
W

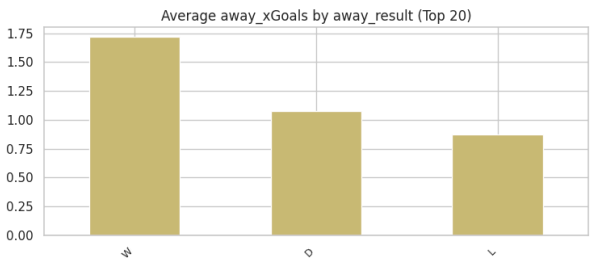

12
10
8
6
4
2
0
W
D
L

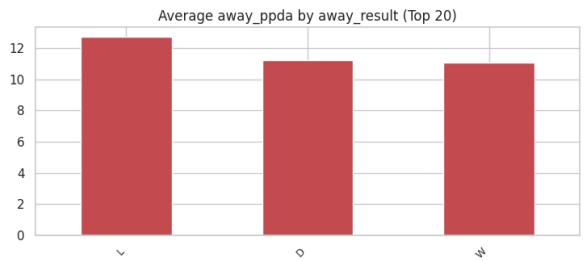

Average home_yellowCards by home_result (Top 20)
2.0
1.5
1.0
0.5
0.0
D
L
W

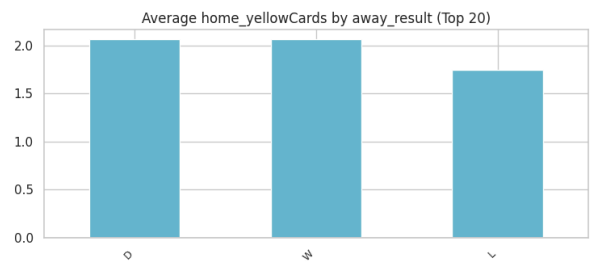

Average away_yellowCards by home_result (Top 20)
2.0
1.5
1.0
0.5
0.0
D
W
L

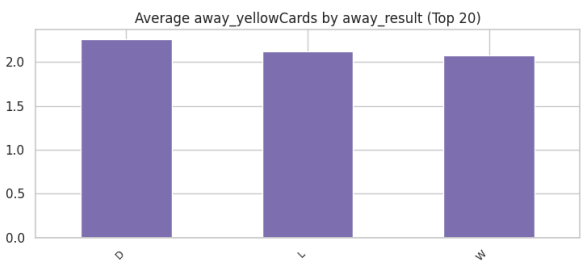

Average home_xGoals by away_result (Top 20)
2.0
1.5
1.0
0.5
0.0
L
D
W

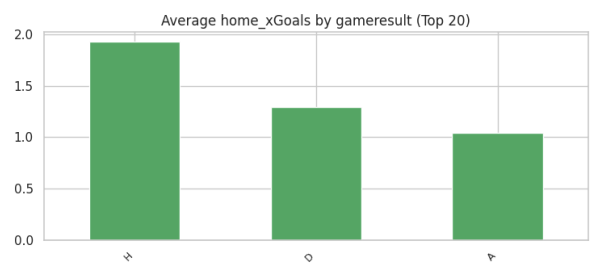

Average home_ppda by away_result (Top 20)
10
8
6
4
2
0
W
L
D

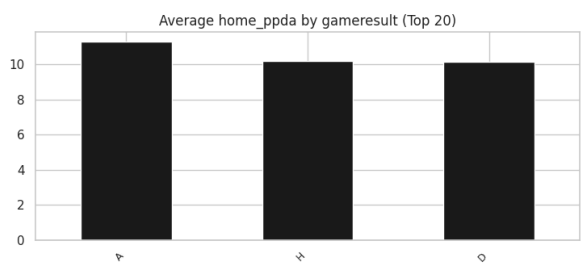

Average away_xGoals by away_result (Top 20)
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
W
D
L

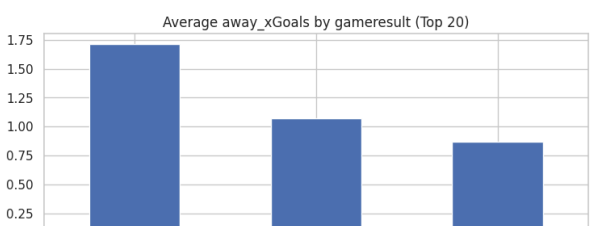

Average away_ppda by away_result (Top 20)
12
10
8
6
4
2
0
L
D
W

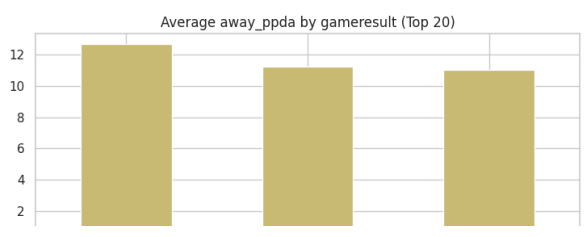

Average home_yellowCards by away_result (Top 20)
2.0
1.5
1.0
0.5
0.0
D
W
L


Average away_yellowCards by away_result (Top 20)
2.0
1.5
1.0
0.5
0.0
D
L
W


Average home_xGoals by gameresult (Top 20)
2.0
1.5
1.0
0.5
0.0
H
D
A

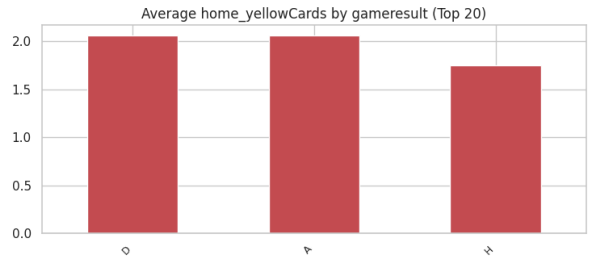

Average home_ppda by gameresult (Top 20)
10
8
6
4
2
0
A
H
D

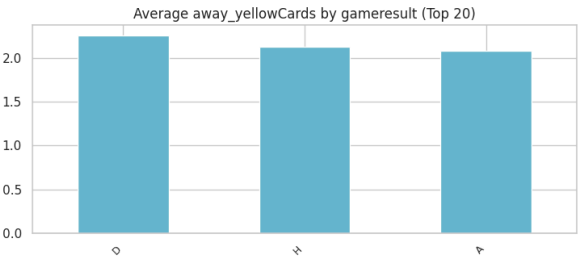

Average away_xGoals by gameresult (Top 20)
1.75
1.50
1.25
1.00
0.75
0.50
0.25

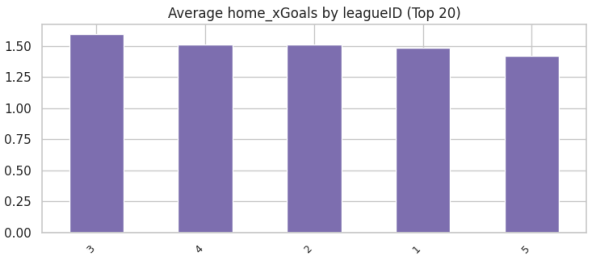

Average away_ppda by gameresult (Top 20)
12
10
8
6
4
2

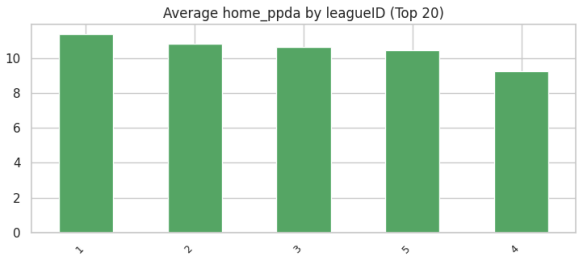

0.00
A
D
H

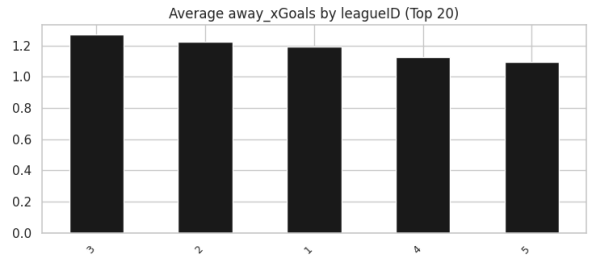

0
H
D
A

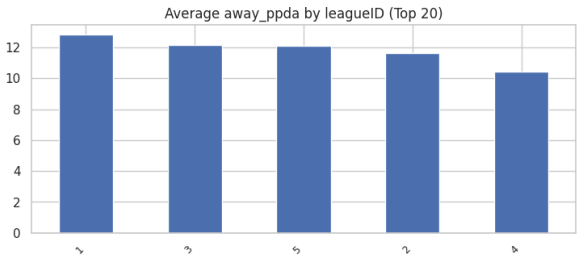

Average home_yellowCards by gameresult (Top 20)
2.0
1.5
1.0
0.5
0.0
D
A
H

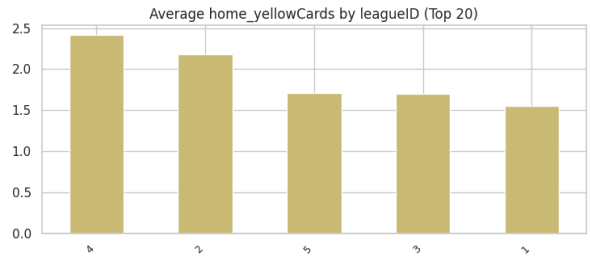

Average away_yellowCards by gameresult (Top 20)
2.0
1.5
1.0
0.5
0.0
D
H
A

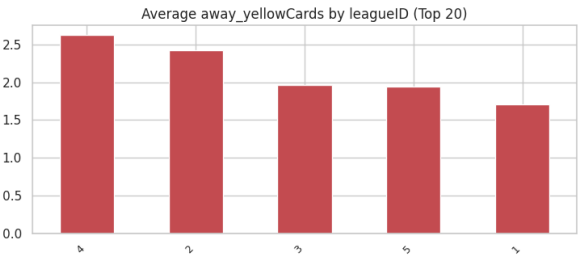

Average home_xGoals by leagueID (Top 20)
1.50
1.25
1.00
0.75
0.50
0.25
0.00
3
4
2
1
5

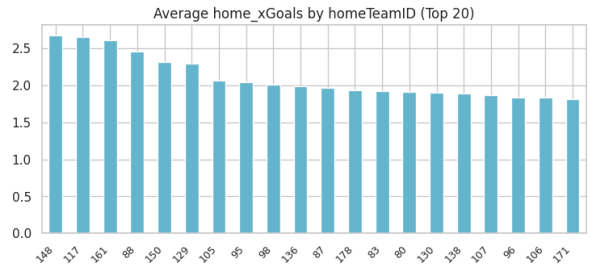

Average home_ppda by leagueID (Top 20)
10
8
6
4
2
0
1
2
3
5
4

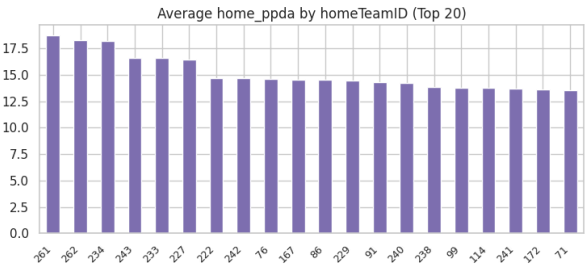

Average away_xGoals by leagueID (Top 20)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
3
2
1
4
5

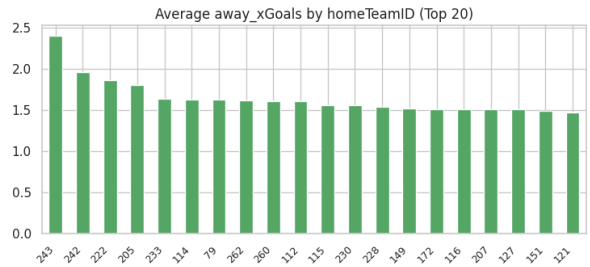

Average away_ppda by leagueID (Top 20)
12
10
8
6
4
2
0
1
3
5
2
4

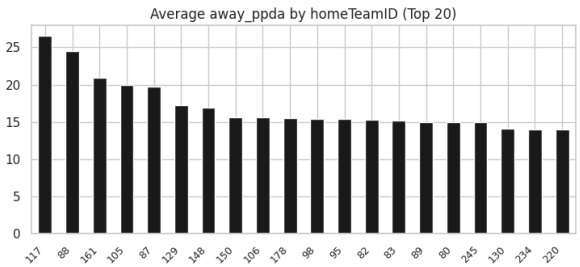

Average home_yellowCards by leagueID (Top 20)
2.5
2.0
1.5
1.0
0.5
0.0
4
2
5
3
1

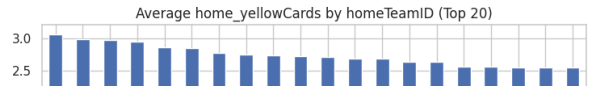

Average away_yellowCards by leagueID (Top 20)
2.5
2.0
1.5
1.0
0.5
0.0
4
2
3
5
1

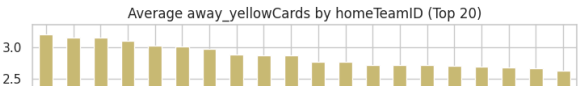

Average home_xGoals by homeTeamID (Top 20)
2.5
2.0
1.5
1.0
0.5
0.0
148
117
161
88
150
129
105
95
98
126
87
178
83
80
130
138
107
96
106
171

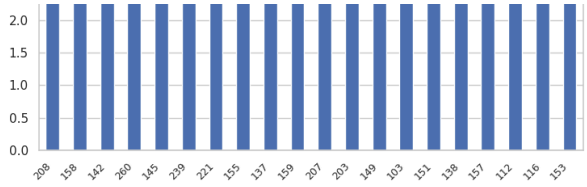

Average home_ppda by homeTeamID (Top 20)
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
261
262
234
243
233
227
222
242
76
167
86
229
91
240
238
99
114
241
172
71

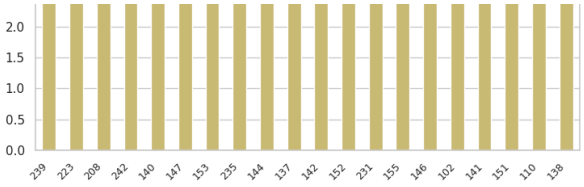

Average away_xGoals by homeTeamID (Top 20)
2.5
2.0
1.5
1.0
0.5
0.0
243
242
222
205
233
114
79
262
260
112
115
230
228
149
172
116
207
127
151
121

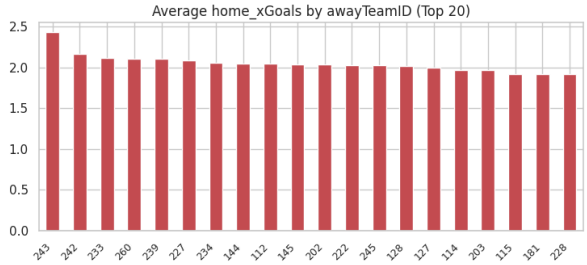

Average away_ppda by homeTeamID (Top 20)
25
20
15
10
5
0
117
88
161
105
87
129
148
150
106
178
98
95
82
83
89
80
245
130
234
220

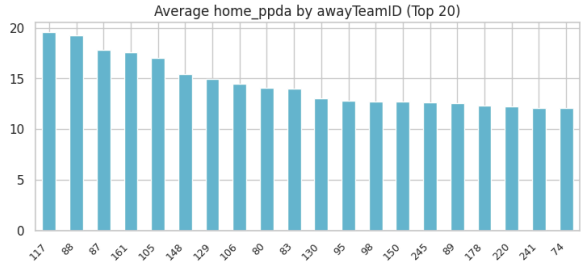

Average home_yellowCards by homeTeamID (Top 20)
3.0
2.5

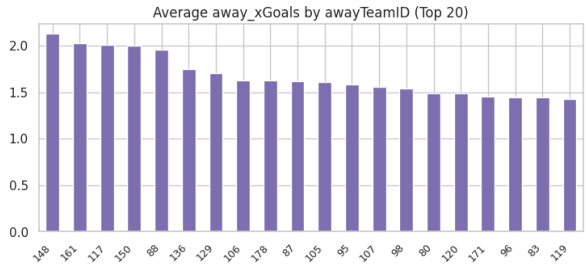

Average away_yellowCards by homeTeamID (Top 20)
3.0
2.5

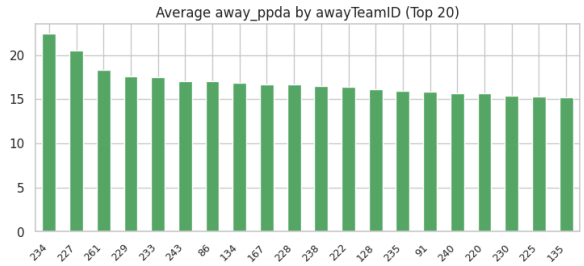


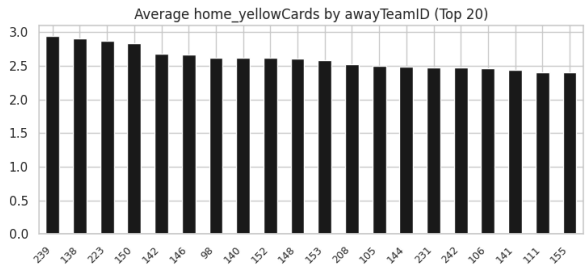


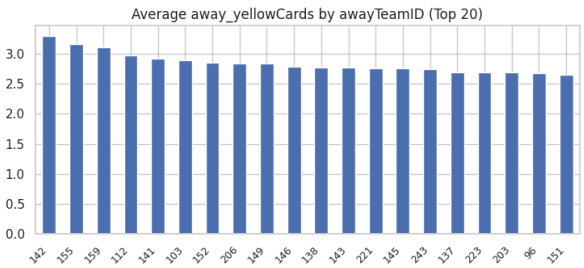

Average home_xGoals by awayTeamID (Top 20)

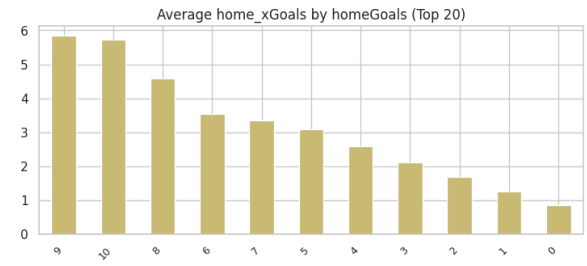

Average home_ppda by awayTeamID (Top 20)

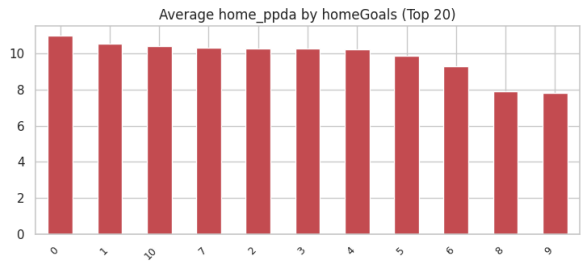

Average away_xGoals by awayTeamID (Top 20)

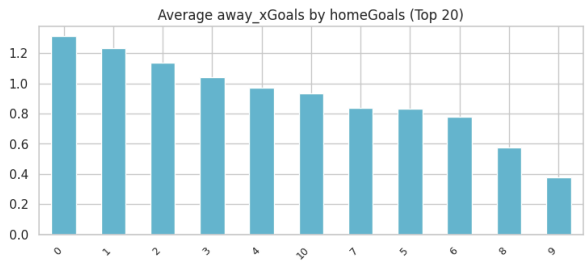

Average away_ppda by awayTeamID (Top 20)

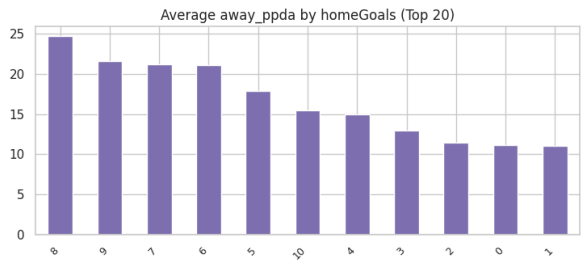

Average home_yellowCards by awayTeamID (Top 20)

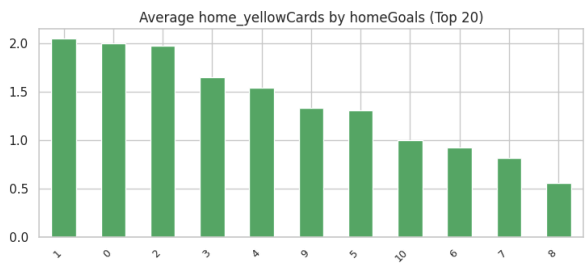

Average away_yellowCards by awayTeamID (Top 20)

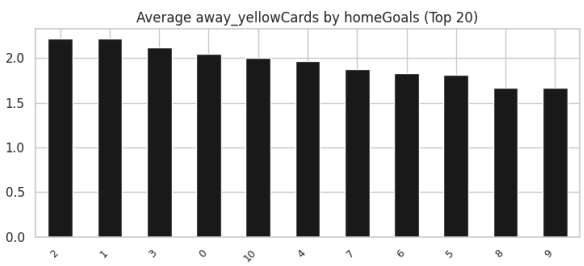

Average home_xGoals by homeGoals (Top 20)

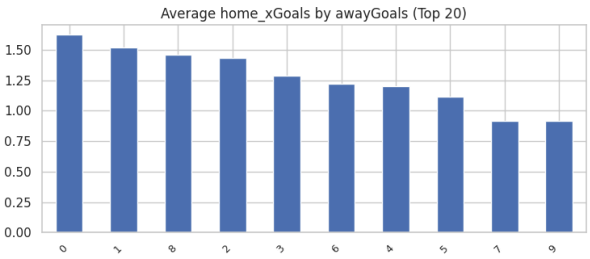

Average home_ppda by homeGoals (Top 20)

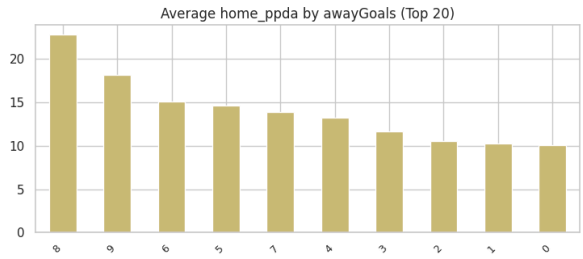

Average away_xGoals by homeGoals (Top 20)

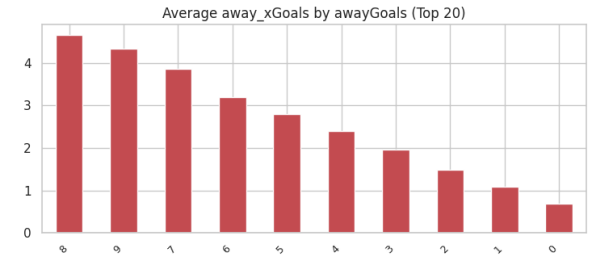

Average away_ppda by homeGoals (Top 20)

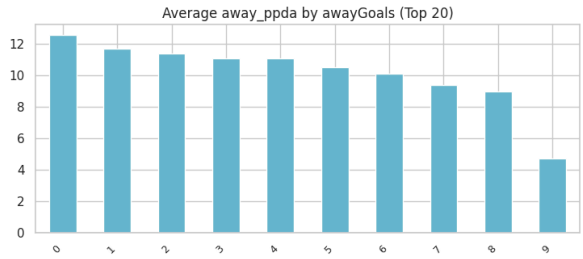

Average home_yellowCards by homeGoals (Top 20)

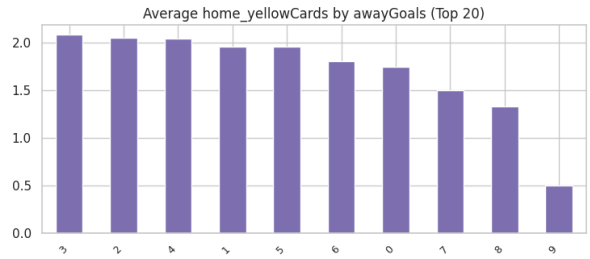

Average away_yellowCards by homeGoals (Top 20)

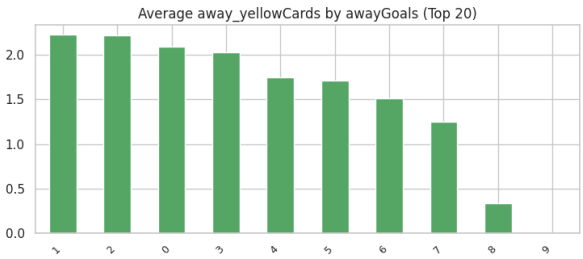

Average home_xGoals by awayGoals (Top 20)

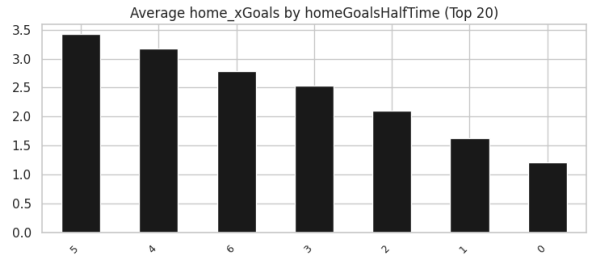

Average home_ppda by awayGoals (Top 20)

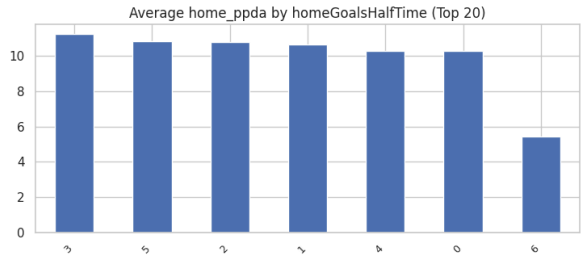

Average away_xGoals by awayGoals (Top 20)

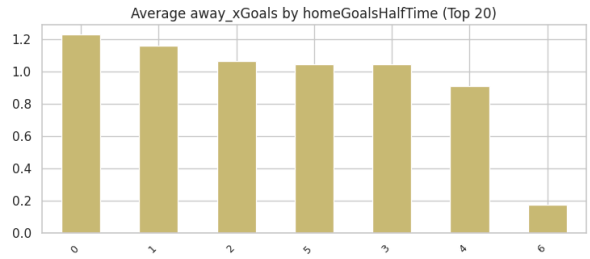

Average away_ppda by awayGoals (Top 20)

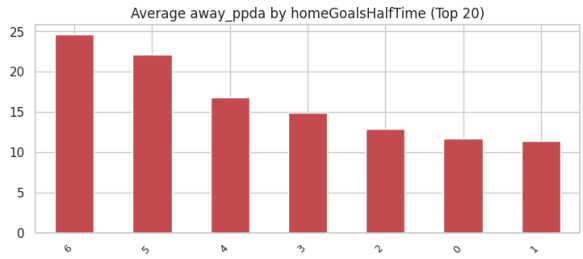

Average home_yellowCards by awayGoals (Top 20)

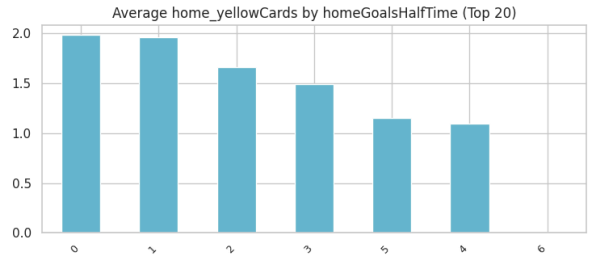

Average away_yellowCards by awayGoals (Top 20)

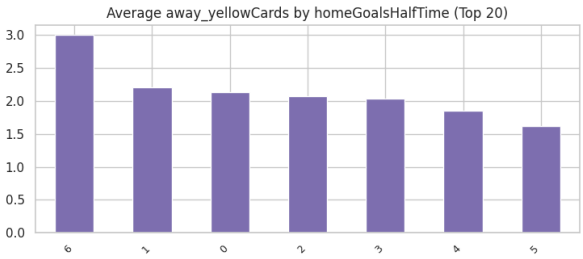

Average home_xGoals by homeGoalsHalfTime (Top 20)

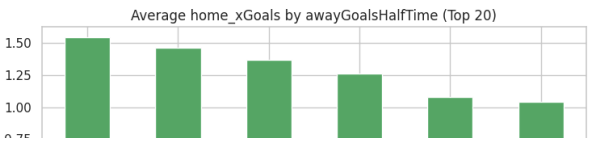

Average home_ppda by homeGoalsHalfTime (Top 20)

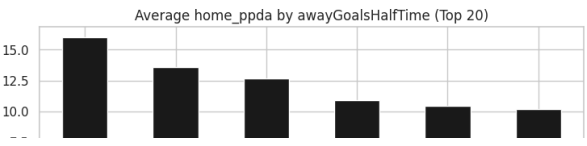

Average away_xGoals by homeGoalsHalfTime (Top 20)

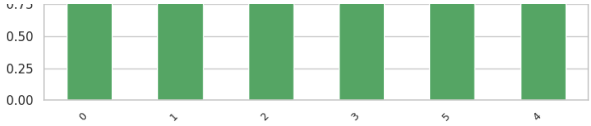

Average away_ppda by homeGoalsHalfTime (Top 20)

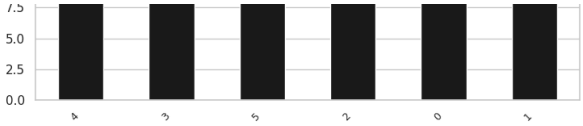

Average home_yellowCards by homeGoalsHalfTime (Top 20)

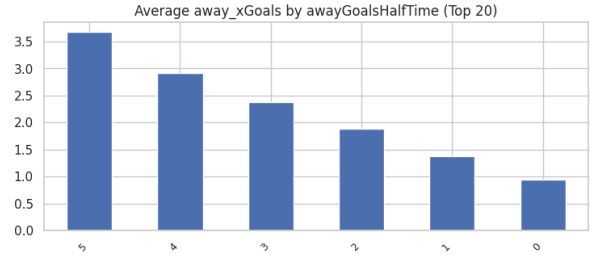

Average away_yellowCards by homeGoalsHalfTime (Top 20)

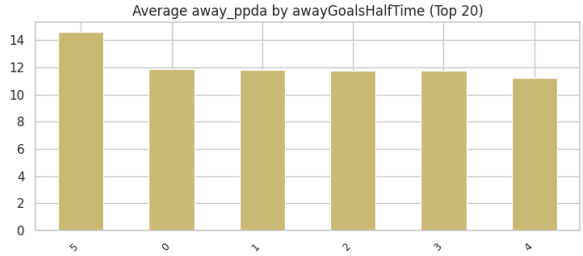

Average home_xGoals by awayGoalsHalfTime (Top 20)

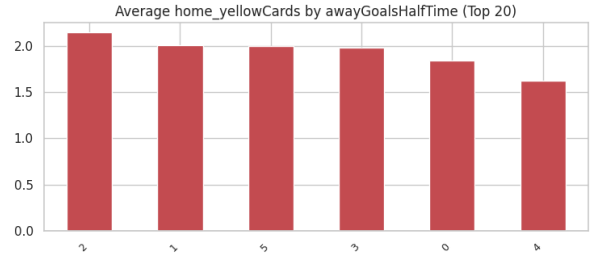

Average home_ppda by awayGoalsHalfTime (Top 20)

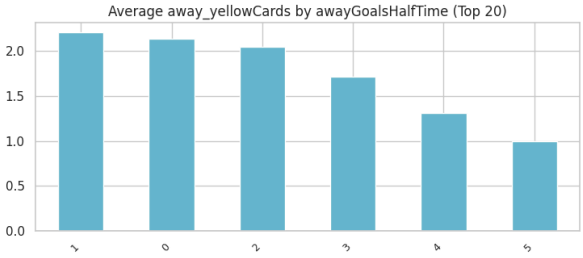


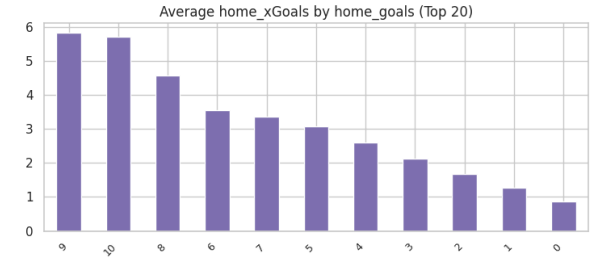


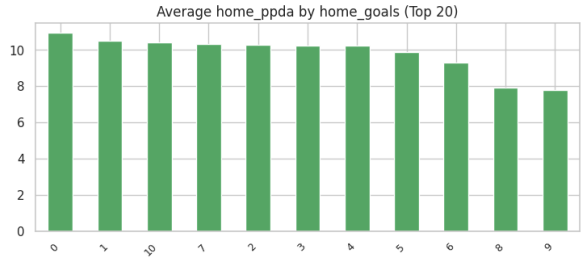

Average away_xGoals by awayGoalsHalfTime (Top 20)

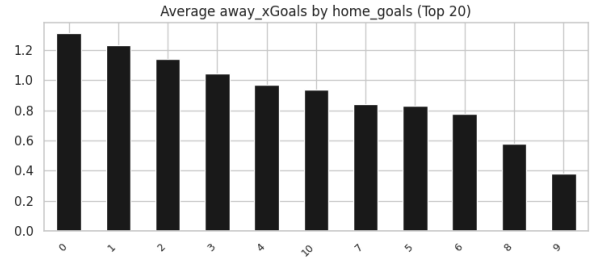

Average away_ppda by awayGoalsHalfTime (Top 20)

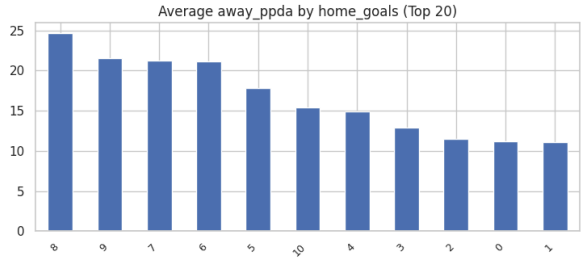

Average home_yellowCards by awayGoalsHalfTime (Top 20)

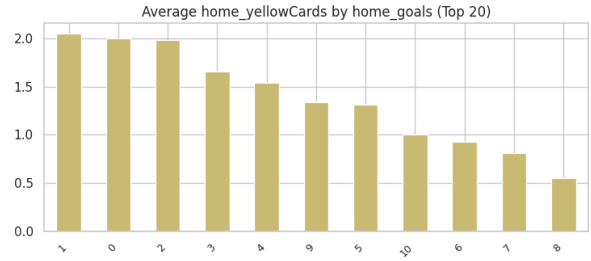

Average away_yellowCards by awayGoalsHalfTime (Top 20)

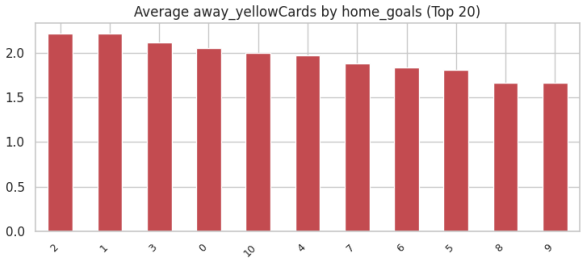

Average home_xGoals by home_goals (Top 20)

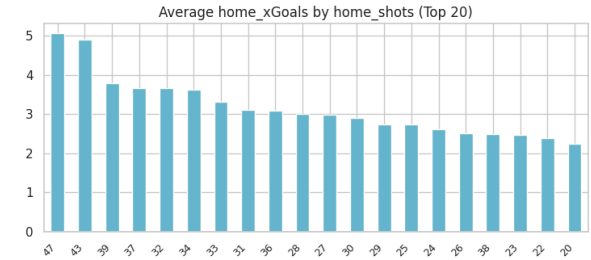

Average home_ppda by home_goals (Top 20)

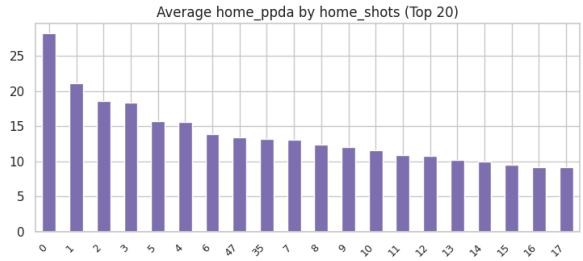

Average away_xGoals by home_goals (Top 20)

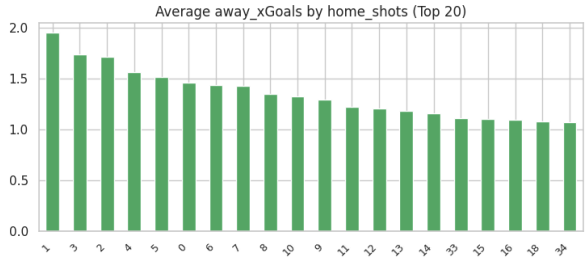

Average away_ppda by home_goals (Top 20)

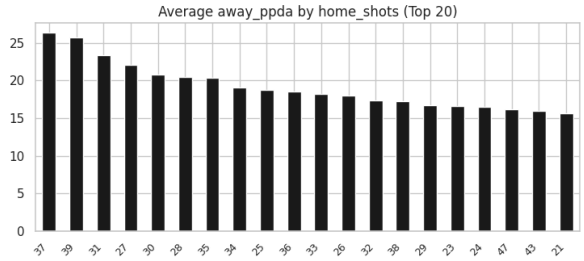

Average home_yellowCards by home_goals (Top 20)

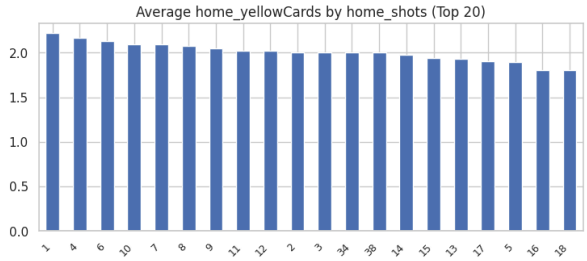

Average away_yellowCards by home_goals (Top 20)

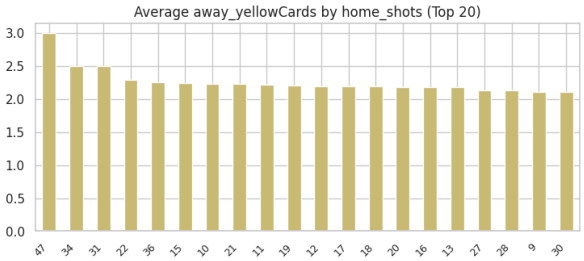

Average home_xGoals by home_shots (Top 20)

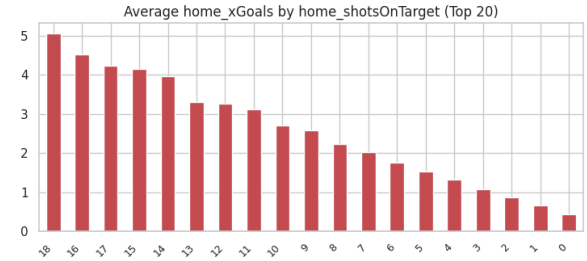

Average home_ppda by home_shots (Top 20)

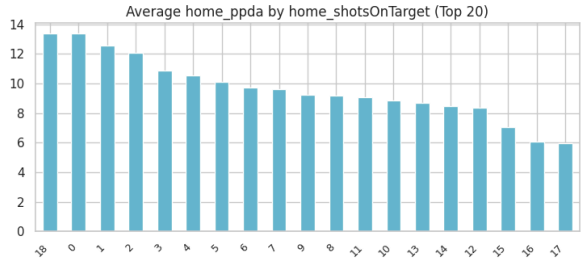

Average away_xGoals by home_shots (Top 20)

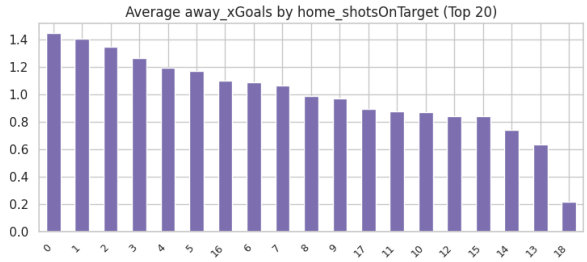

Average away_ppda by home_shots (Top 20)

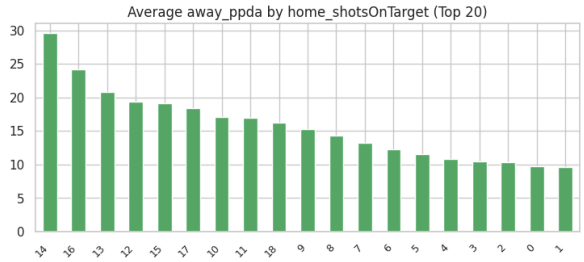

Average home_yellowCards by home_shots (Top 20)

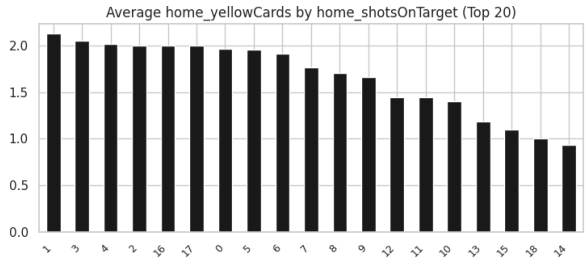

Average away_yellowCards by home_shots (Top 20)

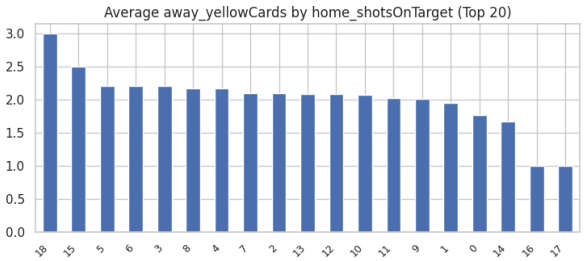

Average home_xGoals by home_shotsOnTarget (Top 20)

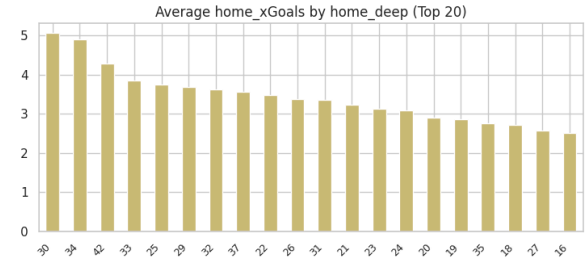

Average home_ppda by home_shotsOnTarget (Top 20)

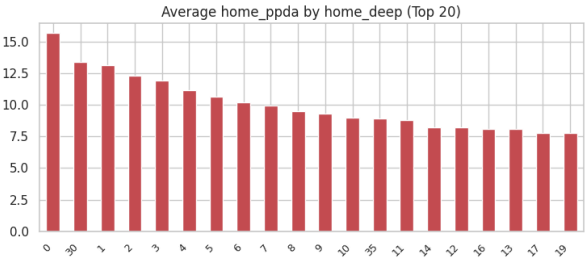

Average away_xGoals by home_shotsOnTarget (Top 20)

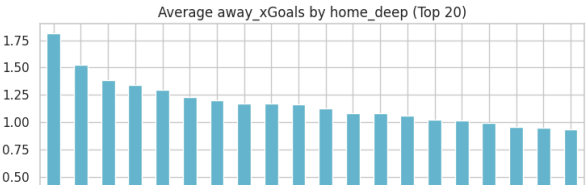

Average away_ppda by home_shotsOnTarget (Top 20)

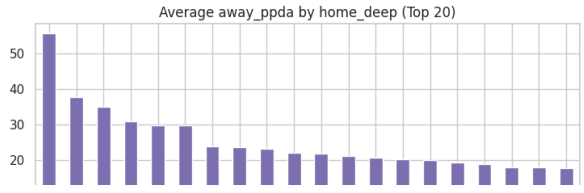

Average home_yellowCards by home_shotsOnTarget (Top 20)

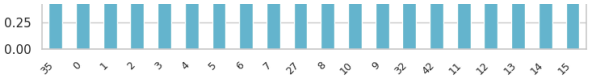

Average away_yellowCards by home_shotsOnTarget (Top 20)

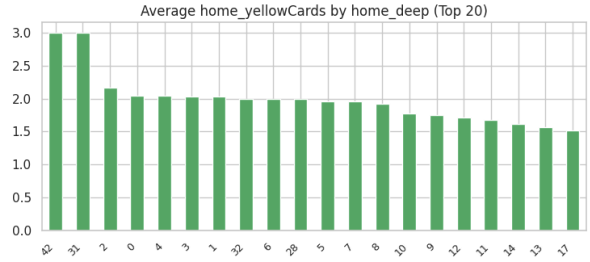

Average home_xGoals by home_deep (Top 20)

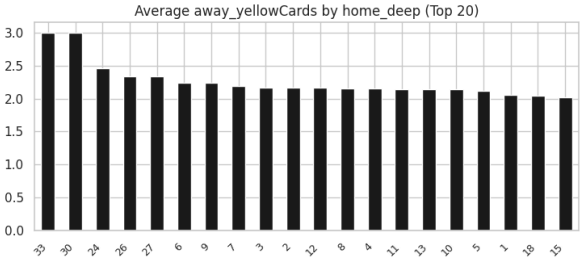

Average home_ppda by home_deep (Top 20)

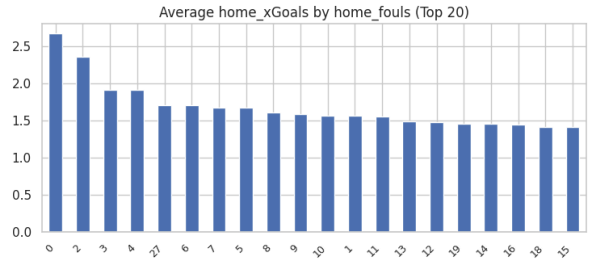

Average away_xGoals by home_deep (Top 20)

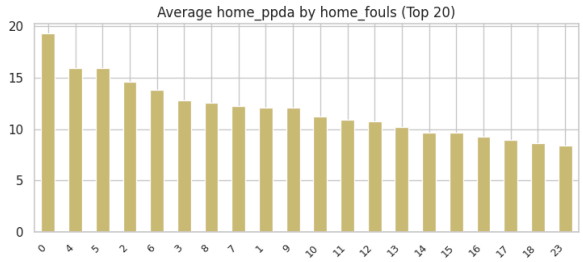

Average away_ppda by home_deep (Top 20)

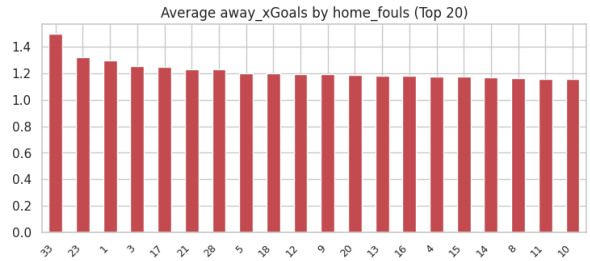


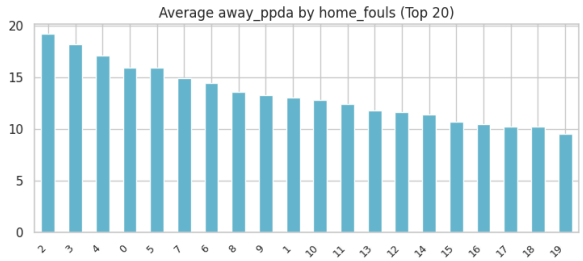

Average home_yellowCards by home_deep (Top 20)

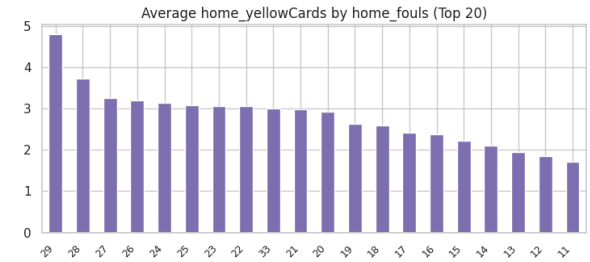

Average away_yellowCards by home_deep (Top 20)

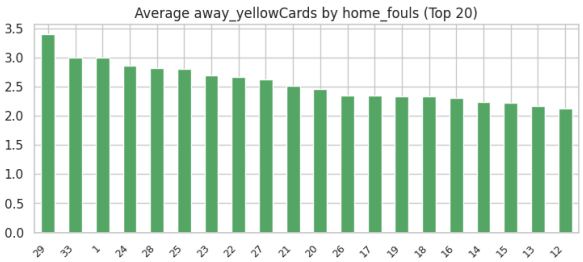

Average home_xGoals by home_fouls (Top 20)

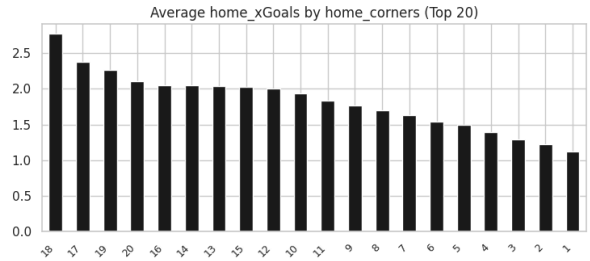

Average home_ppda by home_fouls (Top 20)

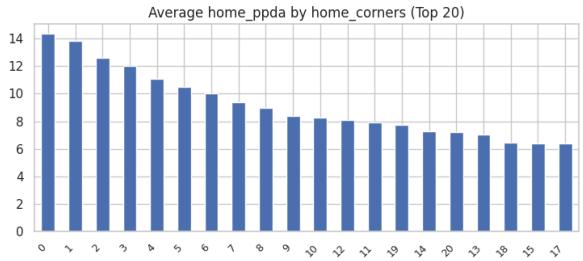

Average away_xGoals by home_fouls (Top 20)

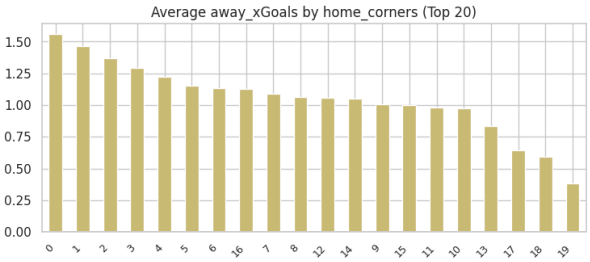

Average away_ppda by home_fouls (Top 20)

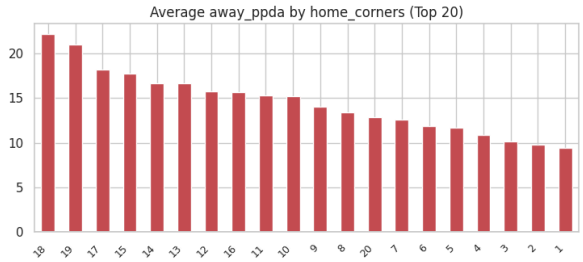

Average home_yellowCards by home_fouls (Top 20)


Average away_yellowCards by home_fouls (Top 20)


Average home_xGoals by home_corners (Top 20)

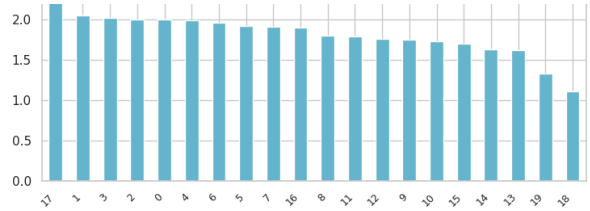

Average home_ppda by home_corners (Top 20)

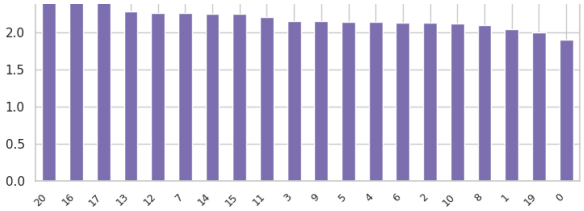

Average away_xGoals by home_corners (Top 20)

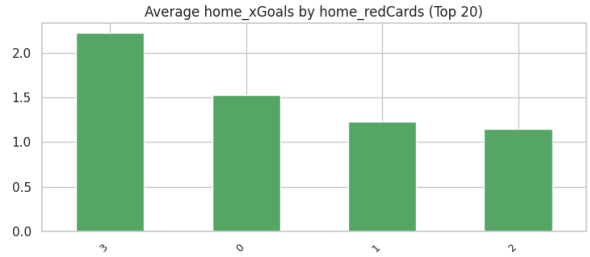

Average away_ppda by home_corners (Top 20)

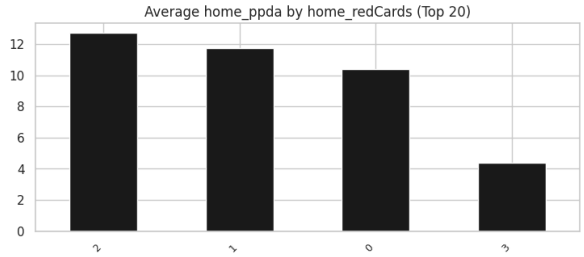

Average home_yellowCards by home_corners (Top 20)

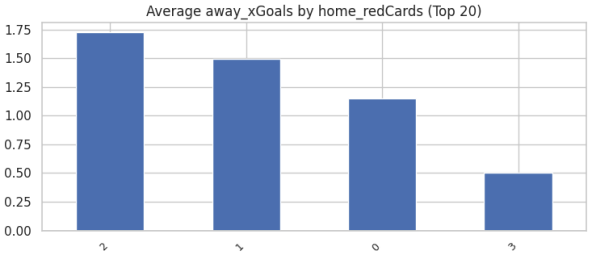

Average away_yellowCards by home_corners (Top 20)

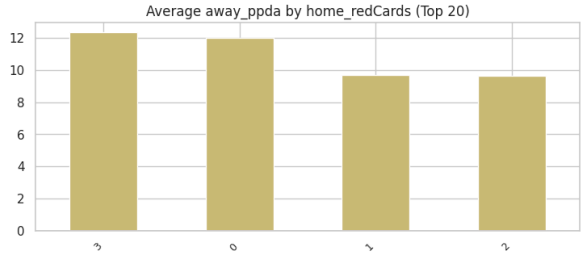


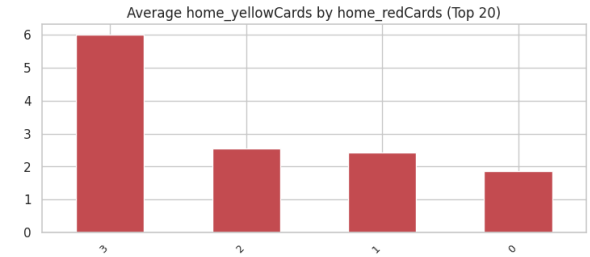


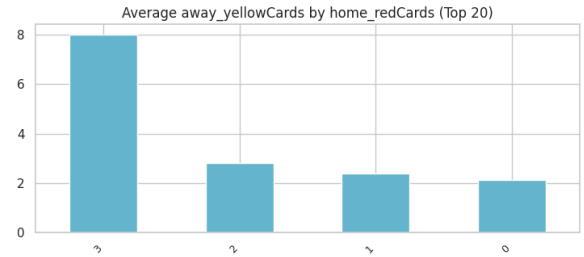

Average home_xGoals by home_redCards (Top 20)

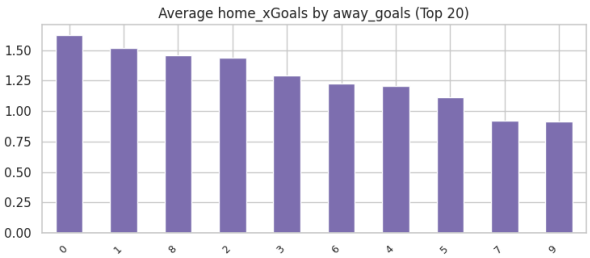

Average home_ppda by home_redCards (Top 20)

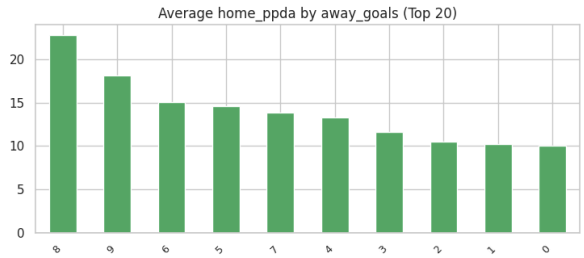

Average away_xGoals by home_redCards (Top 20)

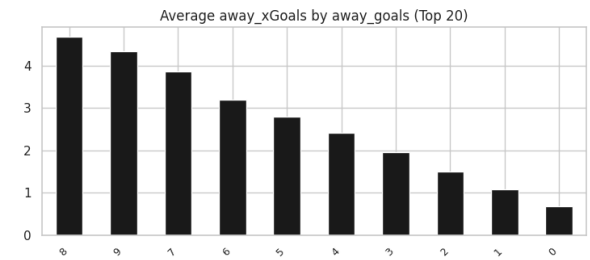

Average away_ppda by home_redCards (Top 20)

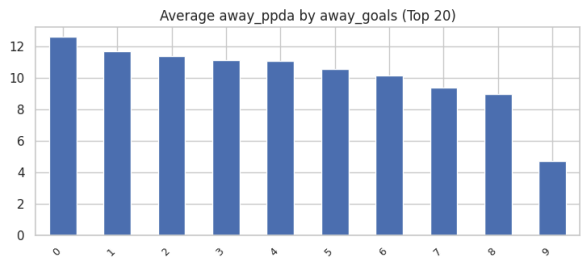

Average home_yellowCards by home_redCards (Top 20)

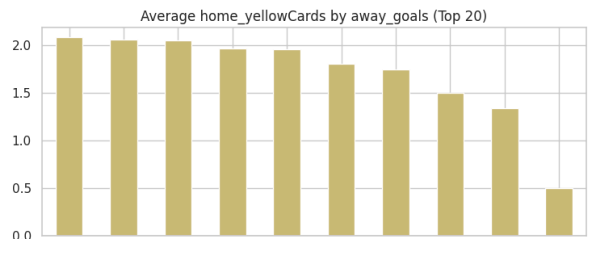

Average away_yellowCards by home_redCards (Top 20)

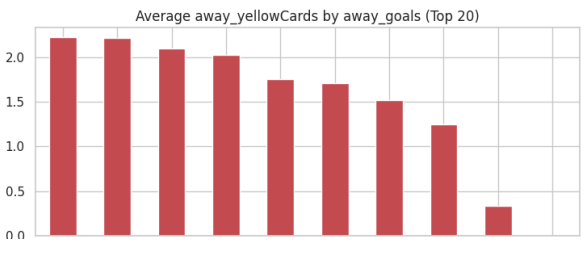

Average home_xGoals by away_goals (Top 20)

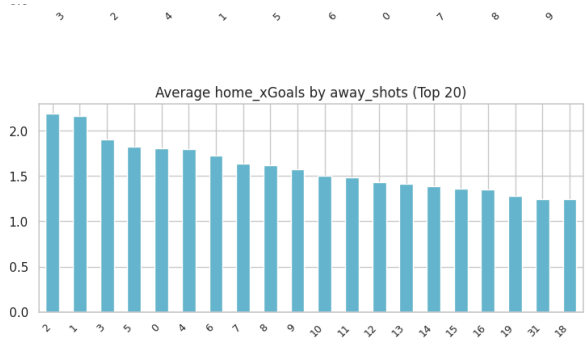

Average home_ppda by away_goals (Top 20)

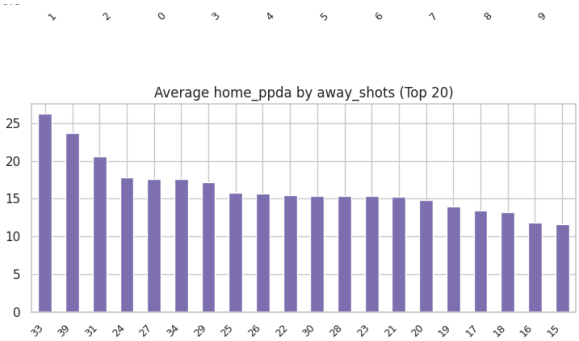

Average away_xGoals by away_goals (Top 20)


Average away_ppda by away_goals (Top 20)


Average home_yellowCards by away_goals (Top 20)

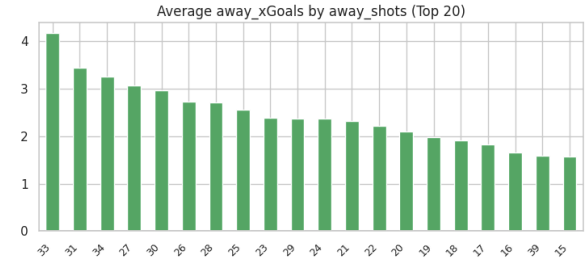

Average away_yellowCards by away_goals (Top 20)

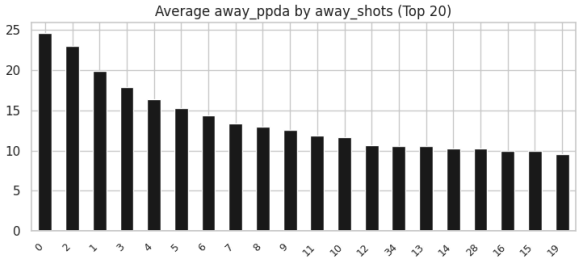


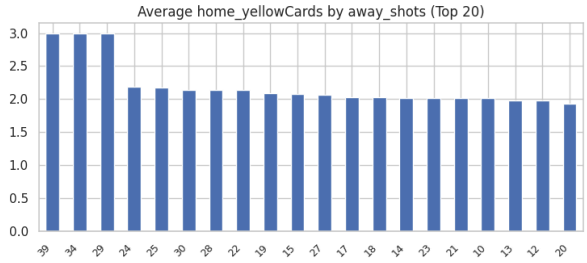


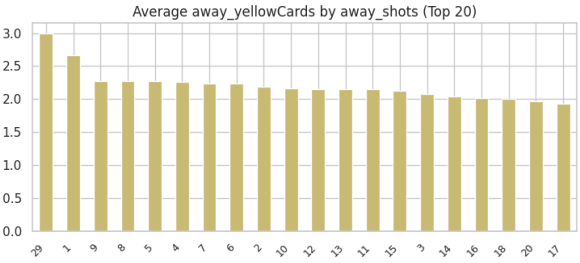
Average home_xGoals by away_shots (Top 20)

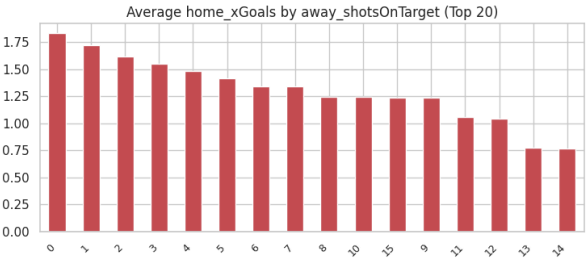
Average home_ppda by away_shots (Top 20)

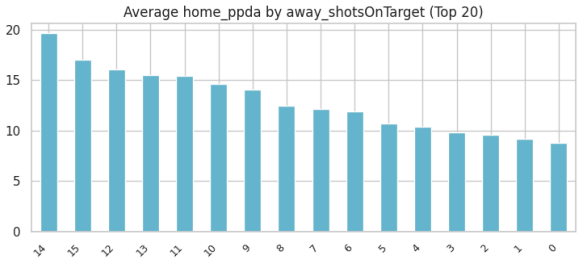

Average away_xGoals by away_shots (Top 20)

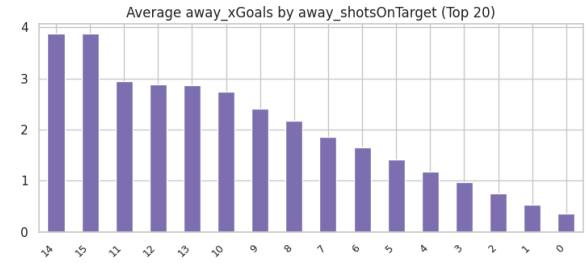

Average away_ppda by away_shots (Top 20)

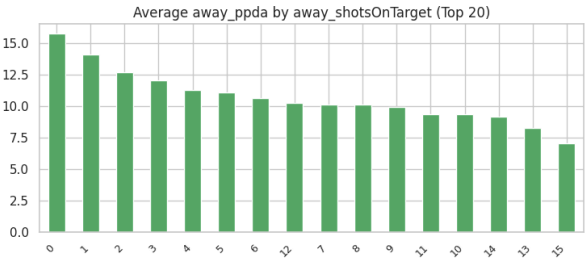

Average home_yellowCards by away_shots (Top 20)

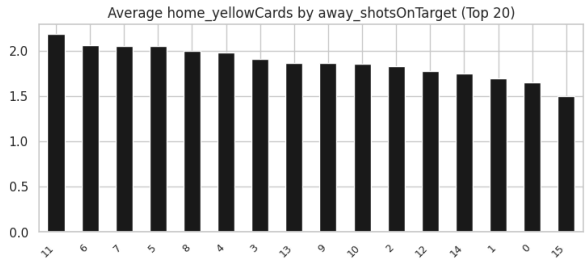

Average away_yellowCards by away_shots (Top 20)

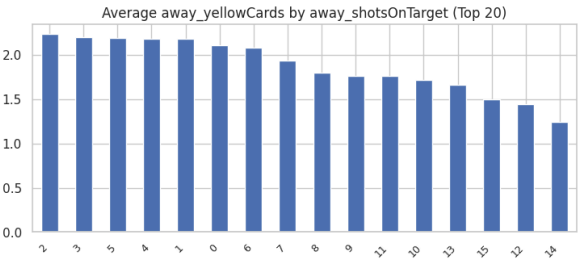

Average home_xGoals by away_shotsOnTarget (Top 20)

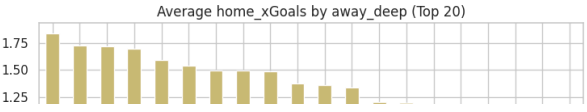

Average home_ppda by away_shotsOnTarget (Top 20)


Average away_xGoals by away_shotsOnTarget (Top 20)

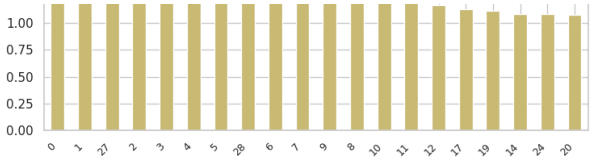

Average away_ppda by away_shotsOnTarget (Top 20)

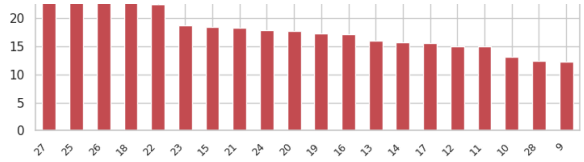

Average home_yellowCards by away_shotsOnTarget (Top 20)

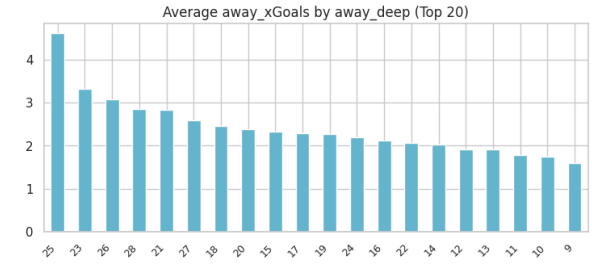

Average away_yellowCards by away_shotsOnTarget (Top 20)

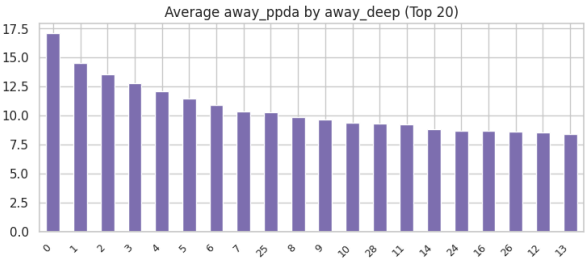

Average home_xGoals by away_deep (Top 20)

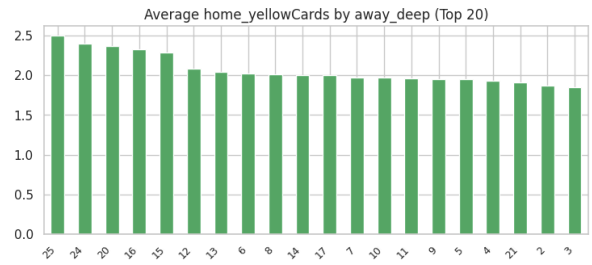

Average home_ppda by away_deep (Top 20)

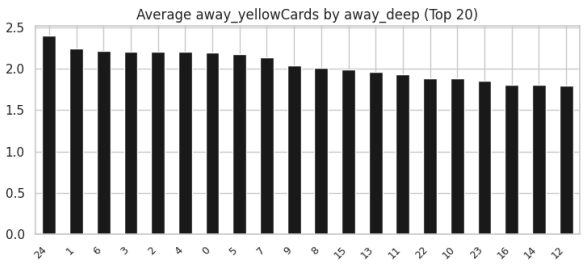


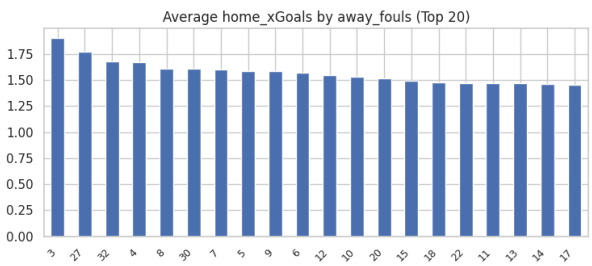


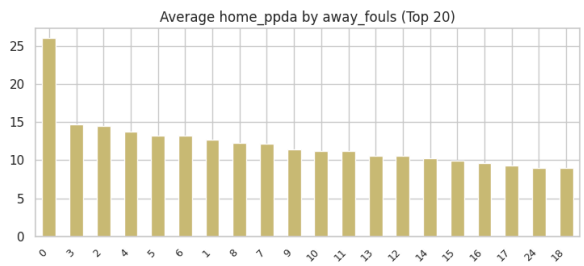

Average away_xGoals by away_deep (Top 20)

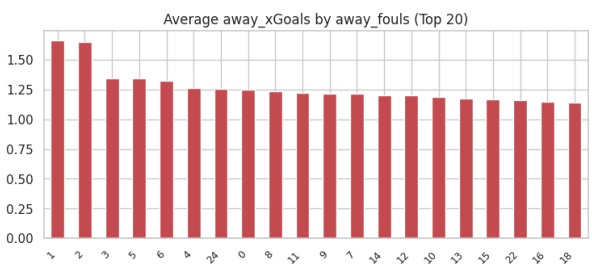

Average away_ppda by away_deep (Top 20)

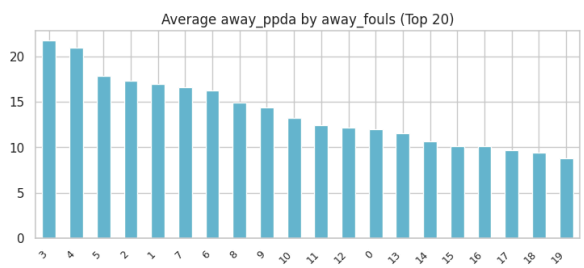

Average home_yellowCards by away_deep (Top 20)

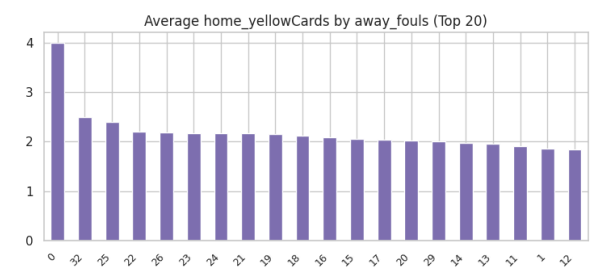

Average away_yellowCards by away_deep (Top 20)

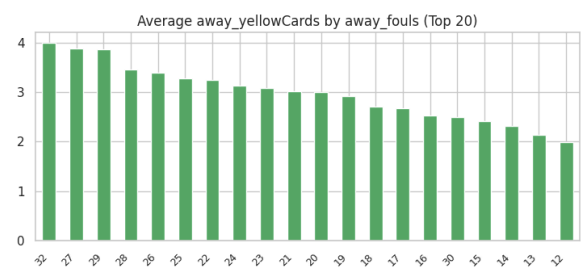

Average home_xGoals by away_fouls (Top 20)

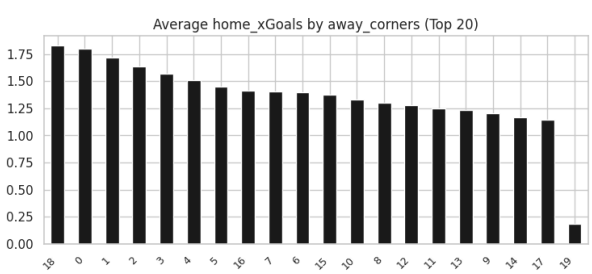

Average home_ppda by away_fouls (Top 20)

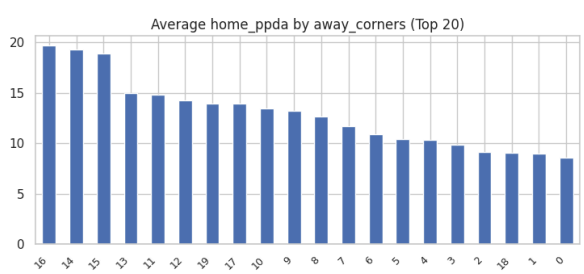

Average away_xGoals by away_fouls (Top 20)

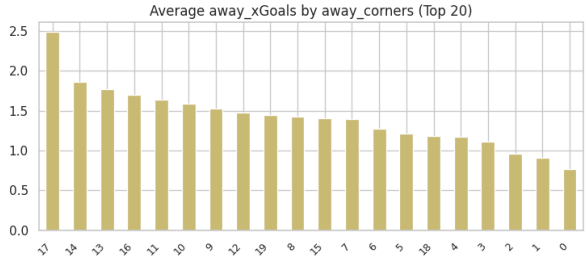

Average away_ppda by away_fouls (Top 20)

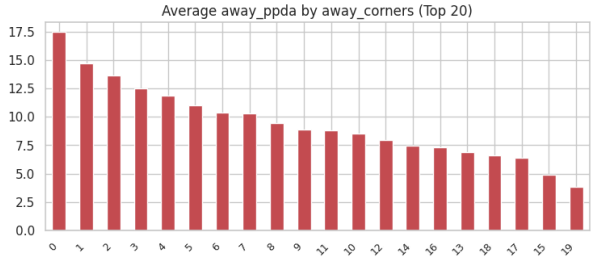

Average home_yellowCards by away_fouls (Top 20)

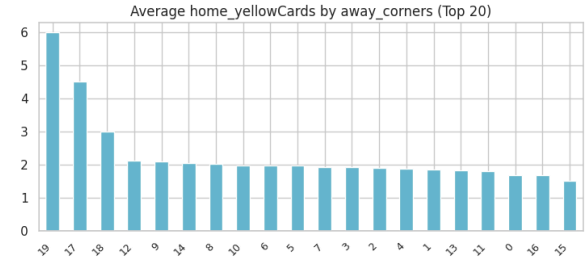

Average away_yellowCards by away_fouls (Top 20)

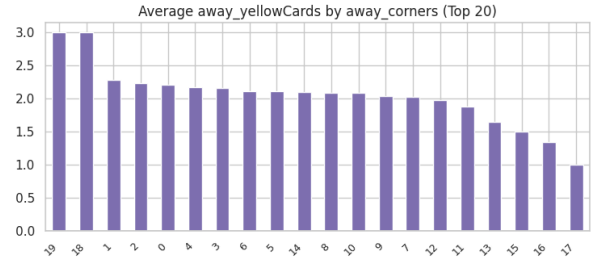

Average home_xGoals by away_corners (Top 20)

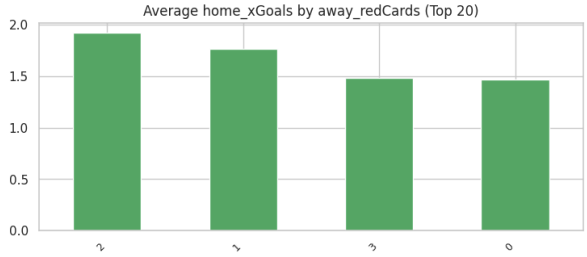

Average home_ppda by away_corners (Top 20)

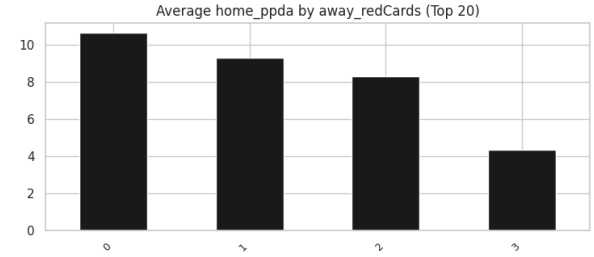

Average away_xGoals by away_corners (Top 20)

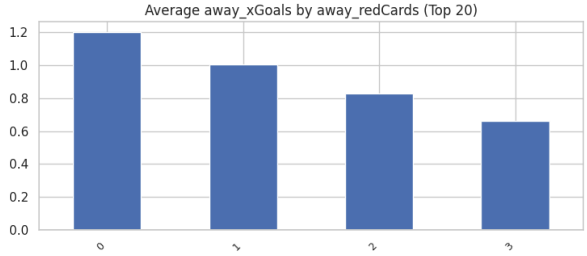

Average away_ppda by away_corners (Top 20)

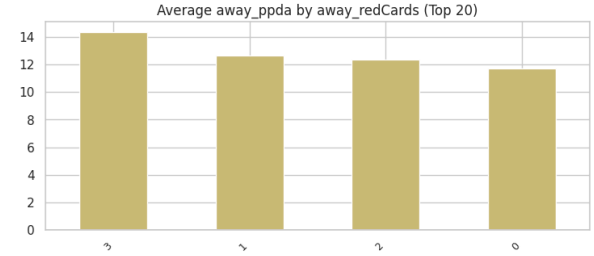

Average home_yellowCards by away_corners (Top 20)

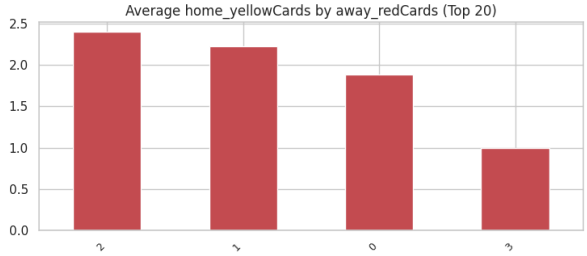

Average away_yellowCards by away_corners (Top 20)

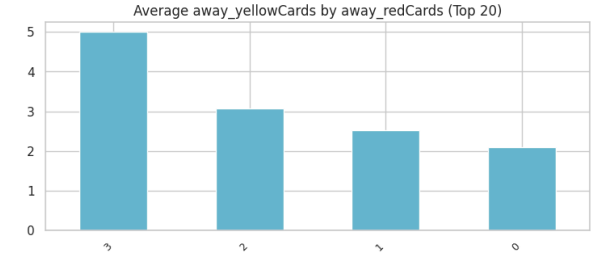

Average home_xGoals by away_redCards (Top 20)


Average home_ppda by away_redCards (Top 20)


Average away_xGoals by away_redCards (Top 20)

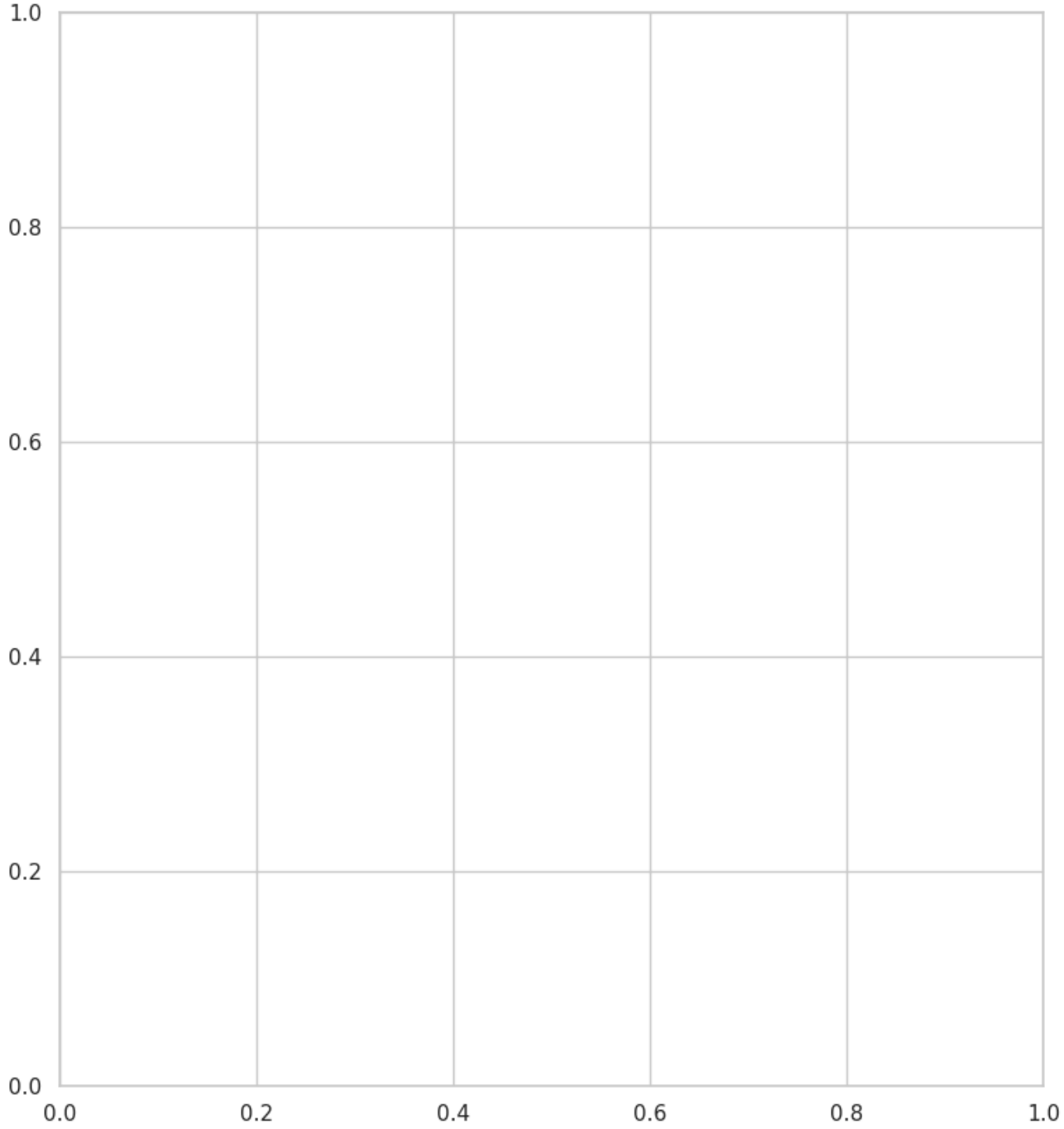

Average away_ppda by away_redCards (Top 20)

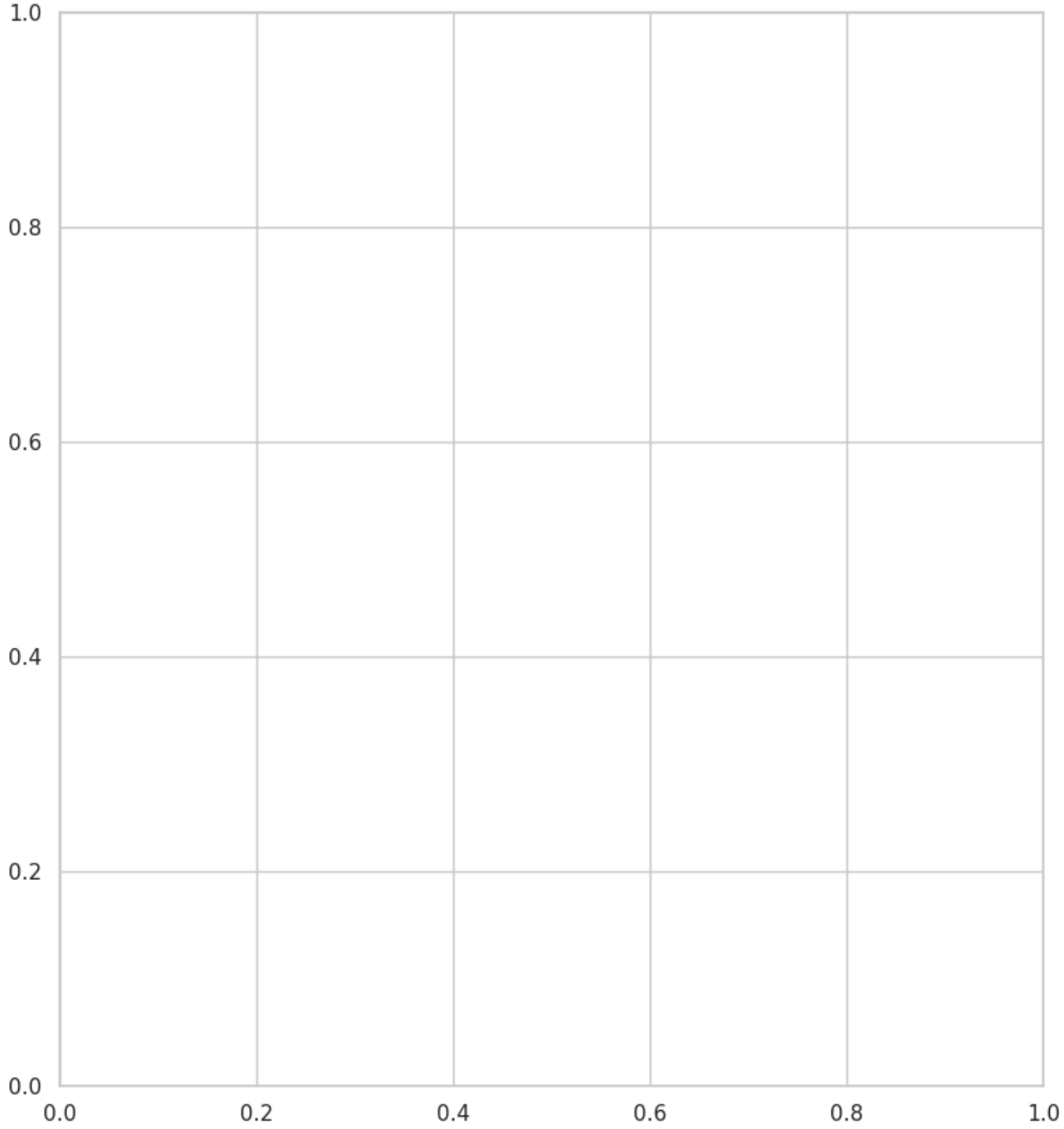

Average home_yellowCards by away_redCards (Top 20)

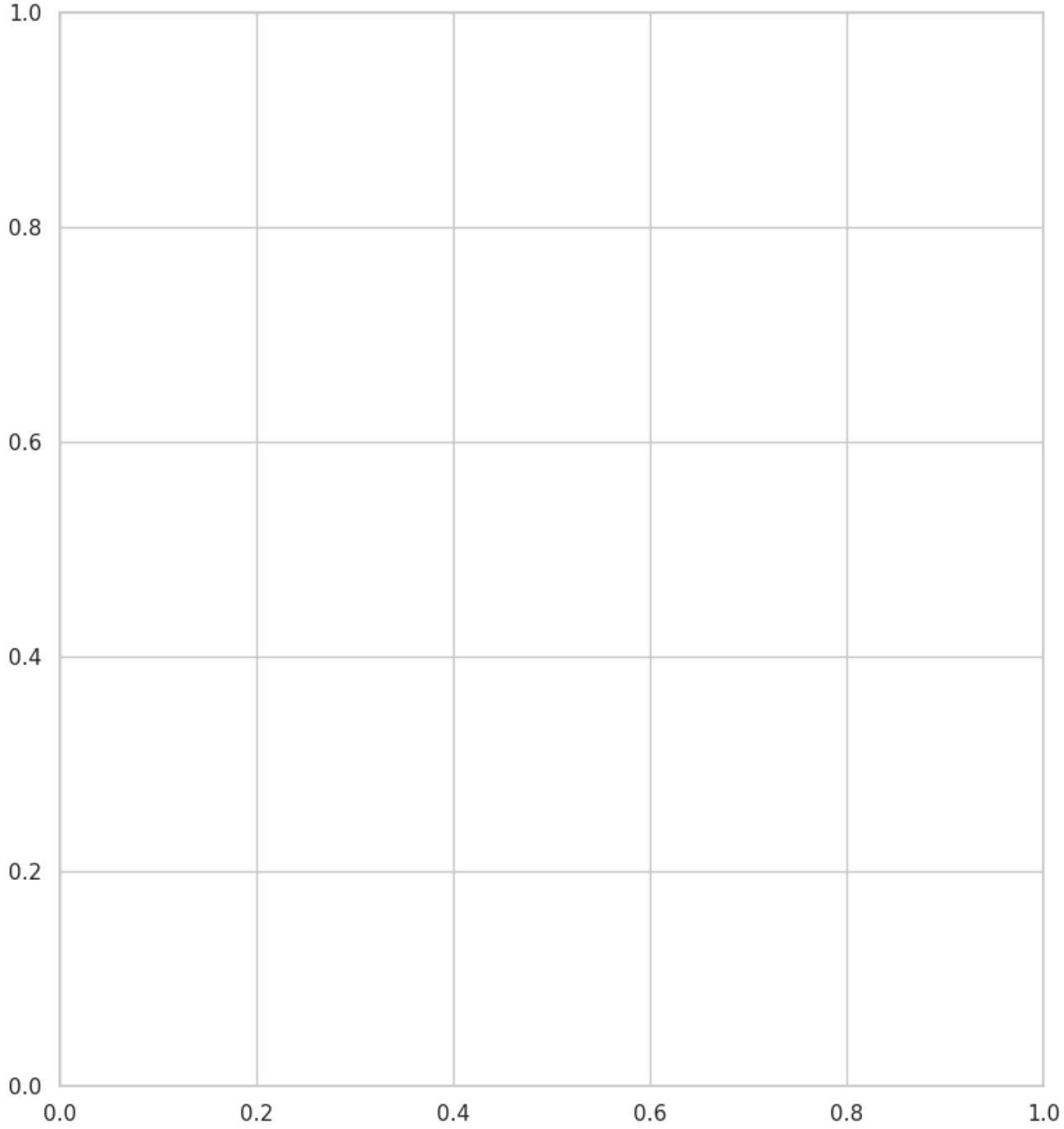

Average away_yellowCards by away_redCards (Top 20)

```
[nltk_data] Downloading collection 'popular'
[nltk_data]    |
[nltk_data]    | Downloading package cmudict to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package cmudict is already up-to-date!
[nltk_data]    | Downloading package gazetteers to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package gazetteers is already up-to-date!
[nltk_data]    | Downloading package genesis to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package genesis is already up-to-date!
[nltk_data]    | Downloading package gutenberg to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package gutenberg is already up-to-date!
[nltk_data]    | Downloading package inaugural to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package inaugural is already up-to-date!
[nltk_data]    | Downloading package movie_reviews to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package movie_reviews is already up-to-date!
[nltk_data]    | Downloading package names to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package names is already up-to-date!
[nltk_data]    | Downloading package shakespeare to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package shakespeare is already up-to-date!
[nltk_data]    | Downloading package stopwords to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package stopwords is already up-to-date!
[nltk_data]    | Downloading package treebank to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package treebank is already up-to-date!
[nltk_data]    | Downloading package twitter_samples to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package twitter_samples is already up-to-date!
[nltk_data]    | Downloading package omw to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package omw is already up-to-date!
[nltk_data]    | Downloading package omw-1.4 to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package omw-1.4 is already up-to-date!
[nltk_data]    | Downloading package wordnet to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package wordnet is already up-to-date!
[nltk_data]    | Downloading package wordnet2021 to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package wordnet2021 is already up-to-date!
[nltk_data]    | Downloading package wordnet31 to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package wordnet31 is already up-to-date!
[nltk_data]    | Downloading package wordnet_ic to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package wordnet_ic is already up-to-date!
[nltk_data]    | Downloading package words to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package words is already up-to-date!
[nltk_data]    | Downloading package maxent_ne_chunker to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package maxent_ne_chunker is already up-to-date!
[nltk_data]    | Downloading package punkt to
```

```
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package punkt is already up-to-date!
[nltk_data]    | Downloading package snowball_data to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package snowball_data is already up-to-date!
[nltk_data]    | Downloading package averaged_perceptron_tagger to
[nltk_data]    |     /home/leoadmin/nltk_data...
[nltk_data]    |   Package averaged_perceptron_tagger is already up-
[nltk_data]    |       to-date!
[nltk_data]    |
[nltk_data]  Done downloading collection popular
Could not draw wordcloud plot for date
Could not draw wordcloud plot for home_date
Could not draw wordcloud plot for away_date
All Plots done
Time to run AutoViz = 54 seconds

 ###################### AUTO VISUALIZATION Completed ########################
=== Completed AutoViz on df_combined ===
```


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0